Representationism and Phenomenism: A critique of two approaches to explaining the relation between representational and phenomenal content

By Hamish David Thompson

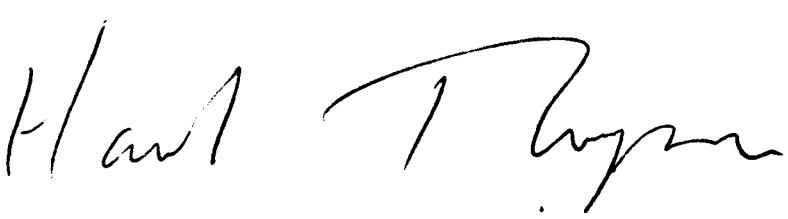
PhD
The University of Edinburgh
2001

I declare the following:

(a) that the thesis has been composed by the candidate, and

(b) either that the work is the candidates own, or, if the candidate has been a member of a research group, that the candidate has made a substantial contribution to the work, such contribution being clearly indicated, and

(c) that the work has not been submitted for any other degree or professional qualification except as specified.



Hamish Thompson

Abstract

The aim of this thesis is to critique two approaches to explaining the relation between representational and phenomenal content. The first approach, representationism, holds that phenomenal content is entirely constituted by intentional or representationalcontent. The second approach, phenomenism, holds that phenomenal content is not entirely constituted by intentional or representational content. There is something 'more' to phenomenal content than just intentional content. Two types of consideration are considered in order to evaluate these two approaches as follows: The first, considers two 'metaphysical speculations', inverted spectra and inverted earth: The second, considers causal and explanatory implications of adopting either of these two approaches. Inverted spectrums and inverted earth provide *prima facie* grounds for rejecting representationism(Block, 1990, 1996; Shoemaker; 1991); however, it will be demonstrated that both 'speculations' do not demonstrate representationismto fail. This thesis will argue that existing responses to inverted earth, for example, Lalor (1999), Lycan (1996), and Tye (1994, 1995b, 1998b) are inadequate. However, it provides a new response to inverted earth on behalf of the wide representationist. Narrow content representationism, which holds that phenomenal contents are functions that map contexts onto contents, will be demonstrated to fail. Doubt is also cast upon teleological approaches to phenomenal content. Causal and explanatory constraints present a problem to both wide representationism and phenomenism. Wide representationalist theories have traditionally faced a challenge from attempting to explain how relational properties can be casually relevant (Fodor, 1987). These problems apply also to representationist theories of phenomenal content. Two current proponents of wide causation are considered, Wilson (1997) and Yablo (1997); their accounts are found to be problematic. Phenomenism either faces a troublesome 'explanatory gap' or a problematic commitment to type-type physicalist identity theories. Finally a proposed future direction for these two theories is suggested.

Acknowledgements

I would like to thank both my examiners (Alexander Bird and Alan Miller) and my supervisors, especially Denis Walsh, for giving my doctoral thesis a new direction. Many thanks also go to my family and friends for their love and support. Thanks also to my brother Matt who very kindly helped print and bind this thesis. A special thanks goes to my wife Heather.

Contents

JACK: You really love me, Gwendolen?
GWENDOLEN: Passionately!
JACK: Darling! You don't know how happy you've made me.
GWENDOLEN: My own Ernest!
JACK: But you don't really mean to say that you would not love me if my name wasn't Ernest?
GWENDOLEN: But your name is Ernest.
JACK: Yes, I know it is. But supposing it was something
else? Do you mean that you couldn't love me then?
GWENDOLEN:(*glibly*): Ah! that is clearly metaphysical speculation, and like most metaphysical speculation has very little reference to the actual facts of life, as we know them.

Oscar Wilde The Importance of Being Earnest

Part I
Inverted Spectrums and Inverted Earth

Chapter 1: Inverted Spectrums

1.1.1 Introduction

This thesis develops and critically evaluates two approaches to explaining the relation between representational and phenomenal content, representationism and phenomenism. Ever since Franz Brentano (1874) described intentionality as the "mark of the mental," scholarship has attempted to explain the mind in terms of intentional content. representationism has emerged and developed out of Brentano's claim about the intentionality of the mental. The intentionality of mental states is a technical way of describing the fact that mental states are directed towards or about something; for example, beliefs, desires and perceptions, are beliefs about 'the weather', desires for 'a cup of tea' and perceptions of 'a tree'. The intentional aspect of experience is characterisedby Robert van Gulick (1995):

> When I have a visual experience of my office laid out before me, my mental state is clearly *about* something or *directed at* something in the intentional sense. My experienceis an experience*of* my office (p.271).

If one maintains that all mental states are necessarily intentional states, then all mental states are never just mental states, but, rather, mental states directed towards some object.[1] Alternatively one might subdivide the class of mental states into distinct categories such as intentional and non-intentional mental states; for example, beliefs desires and perceptions might be classified as intentional while sadness, happiness and pain might be described as non-intentional. A further classification of mental states might be made as follows: *phenomenal* mental states are those mental states that are directly present to our awareness such as a sharp pain; and *non-phenomenal* mental states, for example, a belief about some fact or other

[1] The object that the mental state is directed towards is sometimes described as the intentional or representationalcontent of the mental state.

that might be brought to your attention now, but a few minutes earlier is a belief of which you had no awareness.

It is, however, one thing to describe either some or all mental states as intentional, it is arguably another to claim that the intentional relation between the mental state and its contents *entirely constitutes* the mental state. According to the representationist theory of mind that develops out of Brentano's claim, the intentional relation *is* the mental state; for example, if someone believes that the sun is shining, they possess a mental state, the belief, which is entirely constituted by the 'aboutness' relation between the mental state and the 'sun shining'.

This thesis will develop and evaluate two approaches to *phenomenal* mental states. The first theory, representationism,[2] holds that phenomenal mental states are entirely constituted by their aboutness, their intentional content, or their representational content. The second theory, phenomenism, maintains that phenomenal mental states are not entirely constituted by their aboutness, their intentional content, or their representationalcontent.[3]

Phenomenal mental states appear to have a certain feature that distinguishes them from other non-phenomenal mental states. Phenomenal states are the mental states that are present and at the forefront of awareness. For example, sensations might be described as phenomenal states, because in Nagel's words, "there is something that it is like" to be undergoing that sensation. Van Gulick (1995) expresses the phenomenal aspect to perceptual experience thus:

> My experience has a subjective phenomenal aspect; in Thomas Nagel's evocative phrase, there is "*something that it is like*" to have such an experience. From my first person perspective as the one having or undergoing the experience, my office is present to me in a way that involves a rich array of phenomenal qualities (p.271).

---

[2] Sometimes described as intentionalism or representationalism and examples of proponents of this view are Dretske (1995), Harman (1990), Lycan (1996), Tye (1995a, 1995b, 1998a, 1998b).

[3] Examples of proponents of this view are Block (1990, 1996), Burge(1996), Loar (1990), McGinn (1991), Peacocke (1983).

There is something that it is like to undergo a sharp pain, for example, before one goes into an exam, one undergoes a specific feeling in the pit of one's stomach that has a distinct phenomenal character. The representationist maintains that these phenomenal states are all intentional states; and furthermore, most importantly, the phenomenal content of these mental states is entirely constituted by representational or intentional content.

The second theory, phenomenism, maintains that phenomenal content, or the phenomenal character of mental states is not entirely constituted by intentional content. The phenomenist might argue that there are certain mental states that possess *no* intentional content, but only phenomenal character; one of Ned Block's favourite examples of a non-representational sensation is an orgasm (Block, 1996). Alternatively, the phenomenist might argue that all phenomenal mental states, including some sensations such as fear, are directed at objects so they do possess a phenomenal content, but the intentional content of the mental state does not entirely constitute the phenomenal content.

> I think that sensations - almost always - perhaps even always - have representational content in addition to their phenomenal character. What's more I think that it is often the phenomenal character itself that has the representational content. What I deny is that representational content is all there is to phenomenal character.. . . So the question is better taken as: is there anything mental in experience over and above its representational content? I say yes, the representationist says no (Block, 1996, p.20).

One example that the phenomenist might provide of a mental state that has both a clear intentional content *and* a phenomenal content is the phenomenal content one undergoes when one looks at a ripe tomato. According to the phenomenist, the phenomenal mental states that are directed towards objects are constituted by something more than a mere aboutness relation. They possess a phenomenal content or phenomenal character that goes beyond mere intentional content. According to the phenomenist, mental states that have a phenomenal content are not entirely constituted by intentional or representational content.

1.1.2 Motivations for representationism

One implicit motivation for representationism arises from taking belief states as a paradigm for all mental states. Phenomenal states are merely one subset of all mental states; and thus, one might generalise across all mental states. The claim that intentionality entirely constitutes the mental is a simple, but further, development from noting that intentionality is the "mark of the mental." Therefore, the aboutness relation also entirely constitutes the phenomenal character of these mental states. The representationist, however, clearly requires some argument to carry this line of reasoning through to the conclusion; specifically a justification for why some mental states are phenomenal and others are not.

Another motivation for representationism is to focus on the apparent meaninglessness of the terms that putatively refer to the phenomenal characterof our mental states as only some kind of private intrinsic, non-relational property of our internal mental states to which we have a privileged and intimate access. Ludwig Wittgenstein (1953, 1980) might be interpretedas providing a series of arguments that serve to undermine the meaningfulness of linguistic concepts that refer only to private, inner processes. If someone attempts to claim that their 'pain' refers only to some kind of strongly private inner sensation, this cannot be correct because this inner process would have no outward criteria and the term would be essentially meaningless to another language user -- 'pain' would then mean nothing to another language user.

This line of reasoning perhaps entails a curious form of representationism that maintains that sensation terms and any other concepts describing mental states have the public meanings of these terms when applied by a certain linguistic community. This generates what might be described as 'semantic' representationism.[4] When two language users discuss their mental states and are part of the same linguistic community, their pains, beliefs and desires, refer not to private inner mental states, but to whatever

[4] Rey (1998, p.442) explicitly describes this representationism as 'semantic', and subsequently argues that it fails.

the linguistic community takes these terms to mean.[5] This semantic type of representationism faces some considerable problems when confronted by the arguments of the next chapter.

A further motivation arises from the consideration that as soon as one tries to focus on the phenomenal aspect of one's mental states, one only has access to the intentional content of those mental states. When staring at a patch of red paint, it is not possible to attend to the mere phenomenal character or phenomenal content of the experience because one's attention leaps out into the world.[6] This is sometimes described as the transparency of our mental states (Harman, 1990, Tye, 1995b).

> When you see a tree, you do not experience any features as intrinsic features as intrinsic features of your experience. Look at the tree and try to turn your attention to intrinsic features of your visual experience. I predict that you will find that the only features there to turn your attention to will be features of the presented tree. (Harman, 1990, p.39).

It is arguably one further step to conclude that phenomenal content is entirely constituted by intentional content, but transparency at least offers a *prima facie* motivation for thinking this.[7]

A final motivation arises out of an accusation that the phenomenist commits what is sometimes described as a phenomenological fallacy. The phenomenist when considering the fact that an object looks a certain way, mistakenly identifies the 'look' of something as an independent object that in turn is something of which perceiver is aware (Harman, 1990, p.35). It is not clear that representationism immediately follows from this consideration and arguably the phenomenist can get around this point by claiming that the mental states with phenomenal content nonetheless have intentional content. However, it is not the intentional content that constitutes the phenomenal content, but the *way* that the intentional content

---

[5] Malcolm (1984, pp.49-66) draws on an interpretation of Wittgenstein to justify a rejection of the 'what it is like' or phenomenal character having any reference to an inner private object.
[6] G.E. Moore (1922, p.26) is usually given credit for employing the metaphor of experience as 'transparent' or 'diaphanous'.
[7] Sydney Shoemaker (1991) and Georges Rey (1998, p.441) reply to this argument.

is presented that constitutes the sense in which an object looks a certain way (Block, 1996). According to the phenomenist, it is not the intentional content that entirely constitutes the phenomenal content, but the manner in which the content is presented that constitutes the phenomenal content.

1.1.3 Types of representationism

A semantic style of representationism emerges out of Wittgensteinian considerations, but a variety of other distinct representationisms emerge depending on the specific account of intentional content given by the representationist. Another type of representationism maintains that intentional content of a mental state is determined by the current causal roles that those mental states play in the normal functioning of an organism. This type of representationism is motivated by an attempt to offer a naturalised account of the intentional relation between mental states and objects in terms of the kinds of objects that typically cause those mental states and the kinds of effects that are typically caused by those mental states. (Tye 1995b). This type of representationism might be described as a current long-arm functional representationism for the following reasons: First, the kinds of objects that cause the mental states are usually a part of the environment of the organism, and therefore, the inputs are long-arm and the kinds of effects of the mental states impact on the environment, and thus, the outputs of the mental states are long-arm. Second, the inputs and outputs are the immediate inputs that cause the mental sate and the effects of the mental states are immediate, hence the inputs and outputs are current. Third, the causal roles of the mental states determines the mental states, hence this is a functionalist type of representationism.[8] It is this type of representationism that faces a particularly sharp, although not unanswerable, challenge from the two metaphysical speculations that are considered the next chapters as well as later in this chapter - inverted spectrums and Inverted Earth.

A representationist might not, however, only appeal to *current* functional roles of mental states to determine the intentional content of mental states;

---

[8] Armstrong (1984, pp.169-91) provides a good example of a functionalist account of phenomenal qualities.

the representationist might appeal to the causal roles and the function that the mental states have played in the past to determine the intentional content of those mental states. Different types of historical long-arm functional representationism vary depending on the type of history that is deemed relevant. For example, the representationist might think that it is the manner in which the mental state has developed in individuals within a specific species as that species evolves in its natural habitat that determines the intentional content of those mental states (Lycan, 1996, Tye, 1995b). Alternatively, the representationist might hold that it is the more recent function for which a mental state has been selected as it arises out of a process of learning and conditioning (Dretske, 1981). It is arguably not so much the inverted spectrum argument, which is discussed in this chapter, that is pertinent to the historical type of representationism, but the argument contained within the next two chapters. Historical accounts of intentional content as entirely constituting phenomenal characterare demonstrated to be problematic in chapter three.

A further type of representationism arises out of an attempt to make a distinction between two types of intentional content, wide and narrow (Rey, 1998, Tye, 1994). This account of intentional content is developed partly in this chapter and more completely in next chapter as a response to the metaphysical speculation - Inverted Earth. But narrow contents are ultimately rejected as an adequate account of phenomenal content.

1.1.4 Metaphysical speculations and motivations for phenomenism

One problem that concerned Brentano was how something could be believed if the intentional content or object of the belief does not exist. If the mental state is constituted by the aboutness relation, and one of the relata of the aboutness relation does not exist, how can the aboutness relation, and hence the mental state exist? A belief about no thing is no belief. But we can clearly possess beliefs about non-existent things. This problem led Brentano to speculate that the *apparently* non-existent objects that beliefs can sometimes be directed at are not non-existent, but *in*existent. Whether this reply is successful or not is not a concern of this thesis; however, two

further problems arise out of an appeal to phenomenal mental states being entirely constituted by representationalcontent.

First, it appearsthat mental states can stand in the same intentional relation to the same objects, yet these states might be different with regards to their phenomenal content. The metaphysical speculation concerning inverted spectrums that emerges from this intuition is considered in this chapter.[9] The second problem is that it appears that the intentional content of mental states can change, yet this change is consistent with there being no change the phenomenal content of the mental states. It is this latter problem that forms the grounds for the Inverted Earth argument that is presented against the representationist in the second and third chapters of this thesis.[10]

These metaphysical speculations arguably motivate phenomenism; however, they do not conclusively demonstrate either the falsity of representationism or the truth of phenomenism. They do, however, serve an invaluable heuristic role in both clarifying phenomenism and the distinct types of representationism that emerge as they respond to the inverted spectrum scenarios and Inverted Earth. The overall aim of the first part of this thesis is to examine how some of the above representationisms might be employed as a response to, or are even motivated by, the metaphysical speculations considered in this and the next two chapters. As they are developed further as responses, in chapters two and three, some of these representationisms will be argued to be problematic on grounds independent of the metaphysical speculations. By the end of part one, the argument for phenomenism based on metaphysical speculations will be deemed inconclusive, but only one type of representationism will be deemed to be plausible as an adequate account of phenomenal content or character. In part two, another set of considerations will be developed to help critically evaluate representationism and phenomenism - the casual and explanatory role of phenomenal content.

---

[9] Originally seen in Locke (1680, II, xxxii, 15), more currently examined in Lycan (1973), Shoemaker (1975), Block (1990, 1996, 1999).

[10] Presented by Block (1990, 1996) and defended by Shoemaker (1991) and Rey (1998).

Inverted spectrum scenarios face a host of negative reactions based on downright scepticism that any such fantasies are of any use to doing 'real philosophy.' Or as Gwendolen puts it: "metaphysical speculations have very little reference to the actual facts of life." Reactions to inverted spectrums might be motivated out of a pure wearisome reaction, something like: "oh no, not inverted spectrums again!" Or a patronising "haven't you read your Wittgenstein?" More serious concerns might be grounded on claims that these kinds of arguments are based on assumptions that "beg the question." One's reaction to the scenario depends entirely on the starting assumptions and intuitions. It is perfectly possible to deny the underlying assumptions thus rejecting the supposed conclusions. However, it can be accepted that thought experiments such as inverted spectrums do not serve to conclusively demonstrate the hoped for conclusions, but they are still worth discussing because they are invaluable in helping to clarify the basic assumptions of a theory challenged by the speculation. The reaction to the scenario reveals a plethora of subtle distinctions between positions. To not confront a thought experiment is to lose an opportunity to explicate a theory.

1.2 Inverted Spectrums

Imagine the possibility that two individuals could be looking at the same ripe tomato, yet one individual undergoes a 'red' phenomenal character while looking at the tomato, and the other undergoes a 'green' phenomenal character,while looking at the same tomato. Call 'red' how ripe tomatoes look to *you* and 'green' how unripe tomatoes look to *you*. It therefore really doesn't matter exactly how tomatoes look to you, all that matters is that someone else sees the colour of ripe tomatoes the way the colour of cucumbers look to you. Let us assume this first person has a mental state whilst looking at the tomato with phenomenal character X, and the other person has a mental state with phenomenal character Y. Let us also assume that these two individuals are colour inverted completely around the colour spectrum.[11] These two individuals have mental states that are clearly distinct with regards to their phenomenal content or character.

---

[11] If the two individuals were to look at a traditional colour wheel, each colour would look to the other as the colour on the opposite side of the

What if, however, it is possible that these two individuals have mental states with identical *intentional content*, despite having distinct phenomenal contents or charactersto their mental states? If they have mental states with distinct phenomenal contents or characters, is it possible that they have mental states with identical intentional contents?[12]

This is where the intuitions start to pull apart. What possible scenario could be constructed that would even begin to allow for these individuals to have mental states with identical intentional content? Arguably, the first thing is to assume that the individuals behave in an identical manner. Why does behaving in the same way matter? If intentional contents are determined by the current long-arm functional roles of a mental state, then the way the individuals behave, matters:

> The content of a perceptual representation is functionally defined in part by the ways in which this representation normally arises in perception and in part by the ways in which the representationis used to guide actions (Harman, 1990, p.46).

One long-arm functional role of a mental state is the same as another 'long-arm' functional role of another mental state if both the inputs into the first mental state are the same as the inputs into the second mental state, and the outputs from the first mental state are the same as the outputs of the other mental state. According to this account of content, behaviour matters because it is just one form that the output of mental states takes. If we assume that the two individuals are behaviourally identical (if they behave the same way in the same current environmental context), and their inputs are the same (if they are in the same current environmental context), then we have to assume that the current long-arm functional roles of the two individuals are identical. They have the same input, i.e., when they see the same red tomato, and they have the same output, they interact and respond to the same red tomato in all the same ways. They both exclaim: "look at

---

colour wheel: purples to yellows, reds to greens, all the way around the wheel.

[12] Block argues that inverted spectrums are decisive in demonstrating how the intentional content is the same across the two individuals, but the phenomenal content or character is distinct (1990, pp.53-59).

the tasty red tomato." They both reflect a week later: "how nice the red tomato looked the other day." If both of them believe that they would like to eat a ripe tomato, then the role their perceptual state plays in identifying the tomato as ripe is the same, and combined with their desire for the ripe tomato, they walk over to the ripe tomato, pick it up and eat it. Their perceptual states play identical current 'long-arm' functional roles. The functional roles are wide because the properties of the tomato are part of an external input into the cognitive system of the perceiver. The outputs are wide because the resulting behaviours involve an interaction with objects in the environment. Given an identity in all respects to wide functional roles, we can assume that the two individuals have mental states with identical current long-arm functional roles. [13]

If this current functional identity, combined with a phenomenal difference, seems a little implausible, can the current functional identity between individuals be made any more credible? Arguably the best attempt to make the story more plausible is to consider another variant of the inverted spectrum scenario where the change in phenomenal character occurs within an individual at a certain point in his life, and then the functional role of his mental states can be compared, within the person, before and after the phenomenal change.[14] This is the 'four stage *intra*-personal spectrum inversion' (Block, 1990, pp.60-62):

Stage one, an individual normally interacts with objects and describes ripe tomatoes as "red." Stage two, this individual has colour inverting lenses inserted permanently into his eyes.[15] He is somewhat surprised and starts describing ripe tomatoes as "green." Stage three, he readjusts all his behaviours, including his language, to fit into his linguistic community, so that he starts behaving just as he did before the phenomenal character of his

---

[13] This is how Block formulates an identification of intentional content with long arm functional roles (1990, p.58).

[14] Various versions of intrapersonal spectrum inversions can be seen in the literature such as Gert (1965), Lycan (1973), Shoemaker (1969, 1975), Taylor (1966).

[15] No actual prisms could really be constructed that systematically swap the look of every colour onto its inverted opposite. These lenses have to be assumed to be some kind of mechanism that functionally achieve this task; perhaps they digitally project a changed image onto the retina that swaps every colour onto its inverted opposite: purples to yellows; yellows to purples; reds to greens; green to reds, all the way around the colour wheel.

experiences changed. Stage four, he is given selective amnesia, so that he forgets that there was ever a point where his phenomenal character was switched. This individual behaves in all the same ways as he used to behave before the lenses were inserted. His mental states fulfil all the same functional roles that they fulfilled before the switch, yet the phenomenal content or character of his mental states is switched from before and after the lens insertion.

If the mental states of the individual, compared with stage one and stage four, have identical current 'long-arm' functional roles, then we can assume that the intentional contents of those mental states are identical; however, the phenomenal content or character of the two mental states are different, compared between stage one and stage four. Therefore, phenomenal content or character is distinct from intentional content; representationism is false.

1.3.1 What are the possible representationist replies?

One first reaction might be to allow the possibility of the scenarios, but discount it as unverifiable metaphysical speculation. We live in the real world where these kinds of scenarios are not a "fact of life," but mere exercises in imagination. This reply might become more subtle by distinguishing between types of possibility, whereby the scenario outlined is indeed a 'logical' possibility, but resides in a possible world too far away from our world to be of any practical use. If the possibilities resided in a nomologically possible world 'closer to home,' then the metaphysical speculation might be of some use, but they do not. Daniel Dennett is one philosopher who is extremely unsympathetic to such metaphysical speculation; he claims that there is just no way to verify whether these inverted individuals really are phenomenally inverted. It is in fact constituted in the assumptions of the scenario that it is empirically unverifiable as to whether there are any phenomenal differences between functionally isomorphic individuals (Dennett, 1988).

Dennett, however, misses an opportunity to clarify his theory by offering a more specific reason as to why the phenomenal content or character of the individual does not change if there is genuine functional isomorphism, or

why there is not functional isomorphism between the two individuals. Is the onus on him to reply? No, but it would help clarify his position if he would.

Another denial of the inverted spectrum speculation is motivated out of a reaction that the whole concept 'phenomenal' actually refers to 'nothing' (Malcolm, 1984, Dennett, 1988). This arises out of Wittgensteinian reflections on beetles in boxes (1958, p.101) and the public function of language. These kinds of considerations rely on the fact that sensation terms such as 'pain' or 'phenomenal' do not refer to 'inner objects'. The meaning of such terms is instead determined by the way the term functions in a common language, a language shared by a linguistic community and a shared external environment (Putnam, 1975a, p.175).

If this is correct, then as long as two individuals share a common language within a community and they share a common environment, then those individuals mean the same thing when they express such concepts. If two individuals' colour terms function in the same manner and assuming that they are both part of the same linguistic community and the same environment, then we cannot suppose that one individual when describing the tomato as 'red' could mean something else when another person also describes the tomato as 'red'. If two individuals are part of the same linguistic community and agree with all their colour ascriptions, then inverted spectra are not possible across two individuals.

In the next chapter this 'semantic' style of representationism will be considered further. The next chapter, narrows the representationist field down and assumes that we are dealing with a representationism more sympathetic to at least the concept 'phenomenal character' and that it might refer to 'something.' If it does refer to 'something,' this something is entirely constituted by intentional content. This line is reflected in David Armstrong's reply to Norman Malcolm in which he claims that it is at least an intelligible hypothesis that there may exist something such as 'phenomenal character' and it would be dubious philosophy to rule its existence out just because of considerations motivated by the way language works (Armstrong & Malcolm, 1984, p215-6). When the Inverted Earth

experiment is considered in the next chapter, it is precisely these kinds of anti-private language considerations that motivate one premise of the argument.

Another reaction to the original inverted spectrum scenario is to at least *engage* with the metaphysical speculation, but simply to deny the possibility of functional isomorphism between the two individuals, combined with a difference in phenomenal character. To assume functional isomorphism between the individuals begs the question against the representationist:

> I suggest that [the] inverted spectrum hypothesis will seem evident only to someone who *begins* with the prior assumption that people have an immediate and direct awareness of intrinsic features of their experience (Harman, 1990, p.49).

To even state as a premise in the argument that there could be the possibility of functional isomorphism across the two spectrally inverted individuals begins with the assumption that intentional contents could be distinct from phenomenal character. This is all very well for the *inter*-personal functional isomorphism combined with phenomenal difference, but how about the *intra*-personal spectrum inversion? At what stage does this 'imagined' experiment break down? Harman is probably right, but it is in responding to the argument that the representationist positions might be clarified.

Choices emerge as possible responses to the intra-personal scenario, and different choices might reflect different types of representationism. Something strange might be happening at stage two, and stages three and four appear somewhat tenuous and thus might be taken to be begging the question. One representationist reply might be to offer an account of intentional content not based on current long-arm functional roles, but *historical* long-arm functional roles; therefore, there are factors entirely independent of current functional roles that might be determining content. There are two more interesting lines of defence at this point which might start to differentiate types of representationism depending on how they respond to this scenario as follows: stubborn denial of functional

isomorphism for the mental states at the early (stage one) and later stages (three and four) in the individuals life, or the possibility of phenomenal re-inversion starting at stage two.

1.3.2 Phenomenal re-inversion

If the functional roles of mental states really do change as the individual starts to readjust to his new experiences, perhaps the phenomenal content or character, after the initial switch after the lens insertion, starts to revert back to the original phenomenal content or character.[16] Therefore, before the functional isomorphism is complete, the phenomenal content or character remains distinct at stage two, but as the functional isomorphism begins to establish itself, the phenomenal content or character of the mental states begin to correspond with the earlier states.

> During stage 2 and part of stage 3, interactivity and feedback from the world about the colours of things allows pre-inversion green-detectors to adapt and become red-detectors. At the point in stage 3 at which the informational properties of the perceptual systems reach isomorphism with stage 1, the subject would experience current perceptions as coherent with stage 1's pre-inversion perceptions; both as representing fire trucks as red and clover as green. They would be seen as red and green again. By the time the intentional is back to usual, so is the phenomenal (Lalor, 1999 p.268).

This kind of reply might even be supported by an appeal to real experiments in perception, where individuals wearing spatial inverting lenses, start to phenomenally re-invert (Lalor, 1999, p.269).[17] However, this phenomenal re-inversion occurs in an individual only after the individual undergoes self-motivated gross behavioural readjustment. In actual experiments it was discovered that if an individual wearing spatial inverting spectacles does not interact, in a self-motivated manner, with their environment, then they

---

[16] It is the possibility of phenomenal re-inversion that Dennett takes to be unverifiably indistinguishable from phenomenal change with identical behaviour that leads Dennett to conclude that the concept qualia is meaningless (1988).

[17] Spatially inverting lenses are lenses that left-right re-orient the look of the world such that things that would normally look to be on your left would look to be on your right; and things that would look to be on your right would in fact be on your left. For example: if you were wearing these lenses, you might try to reach out for an object to your left but to your surprise your hand would go out to the right.

undergo no adaptation. Experimental subjects wearing spatially inverting spectacles were split into two groups: one group were wheeled around in carts; they underwent no behavioural adaptation; however, the individuals who could actively interact with their environments by their own motivation adapted (Held & Blossom, 1961).

This reply to inverted spectrums based on phenomenal re-inversion involves an explicit commitment to a wide representationism that relies on the current long-arm functional roles of mental states determining content. Here we have, as a response to inverted spectrums a more clearly defined representationism. Does the inverted spectrum show that this position is false? No, however, when another thought experiment is considered in the next chapter, Inverted Earth might well present a challenge to this theory if this theory appeals to phenomenal re-inversion justified by empirical experiments. It will be argued that the empirical evidence drawn on to support phenomenal re-inversion when applied to these inverted spectrum scenarios is not applicable to the Inverted Earth scenario. Additionally, further grounds will be presented as to why phenomenal re-inversion is problematic in the Inverted Earth scenario. However, there are additional representationist replies to intra-personal spectrum inversions.

1.3.3.1 Denial of functional isomorphism

The representationist may attempt to reply to the intra-personal scenario by assuming that the phenomenal character does change while at the same time denying that there is current functional isomorphism between the individual at stage one, and either stage three or stage four. This reply might be motivated by the mere assumption of representationism (cf. Harman, 1990), or this denial of current functional isomorphism might be given some justification.

1.3.3.2 Fine-grained functional differences

If there are fine-grained functional differences between the individual at stage one, and the individual at stage four, some of their subtler behaviours might not be the same.

> In each case, a range of coarse-grained functional isomorphism is assumed, from which it is concluded that overall there is complete functional identity. Patently, however, there can still be some salient fine-grained functional differences (Tye, 1995b, p.202).

Michael Tye describes these fine-grained functional differences as 'narrow'; this terminology might lead to a serious misunderstanding as to what kind of account of content he is attempting to give, especially given his earlier adherence to a 'narrow' account of content in an earlier paper (Tye, 1994); therefore, this reply might be described better as an appeal to lingering fine-grained functional differences that superficially do not appear when considering the latter stages of the inter-personal inverted spectrum scenario. Incidentally, if only broad functional differences are considered, fine-grained differences do not emerge, despite their presence, also in *inter*-personal inverted spectrum scenarios.

The representationist argues, as part of this response, that an essential premise of the intra-personal spectrum argument is that there are absolutely no behavioural differences, as a result of the mental states, between the individual at an earlier and later stages in his life, otherwise his outputs might be different. According to this representationist response, if *any* differences arise concerning the outputs of mental states, then these mental states have different contents. This complete functional isomorphism requires that the individual at the early stage and at the later stage have coextensive discriminability:

> Jack and Jill must have coextensive colour discriminations and judgements, so that for any two stimuli x and y, Jack finds x indiscriminable from y if and only if Jill does as well. Jack and Jill must also make precisely the same judgements of relative similarity, so that Jack judges x to be more similar to y than to z if and only if Jill does as well (Clark, 1985, p.431).

However, given the way our perceptual systems actually work, there might still remain fine-grained functional differences that do not appear on a superficial inspection of this individual's behaviour at stage four. Gwendolen might also be pleased by the attempted justification for this line

of response to inverted spectrums because it is also grounded by "facts of life."

The representationist who appeals to the relevance of fine-grained functional differences also assumes that the type of spectral inversions in these scenarios are complete colour inversions around the colour space; for example, purples go to yellows; greens go to reds. These complete spectral inversions are then problematic for functional isomorphism, if our colour spaces are asymmetrical. Given an asymmetric colour space, assuming a complete inversion, coextensive discriminability will be lost across the two individuals.

> Interestingly, the need for coextensive discriminations shows that the supposition of spectrum inversion is contrary to fact. Between any two individuals there is likely to be some part of the spectrum for which one of the individuals has better discriminations than the other, so that there are at least two colours x and y such that one individual can discriminate x from y while the other cannot (Clark, 1985, p.432).

Our colour spaces are asymmetrical. What this means is that we are able to make finer discriminations between certain ranges of colours depending on the area of the colour space; furthermore, no symmetrical inversion, through an axis, or even a point rotation, could possibly be constructed, given the 'shape' of our colour space, which would maintain coextensive discriminability (Clark, 1985). Therefore, these fine-grained functional differences across our two individuals can be verified by testing discriminable abilities between individuals. For example, suppose, we are able to identify more shades of purple, than of yellow. What does this imply for an individual who undergoes a colour inversion, from purple to yellow? Subtle fine-grained behavioural differences between this individual, at an early stage, and at a later stage show up. Where once our individual was able to distinguish between many types of sky, by the time he reaches stage two of Block's process, he is not. By the time he reaches stage four, he is *still* not able to discriminate between the types of blue sky that he used to. Where once our individual was only able to discriminate between limited shades of ripe bananas, by the time he reaches stage two, he is able to discriminate between many *more* shades of ripe bananas. By the time he

reaches stage four, he is *still* able to discriminate between many more shades of ripe bananas.

However, what if these kinds of fine-grainedfunctional differences are really *not* relevant to content differences? Tye (1995b) just assumes that fine-grained functional differences are salient to the representationist. Block (1999), on the contrary, argues that fine-grained functional differences arising from a failure of coextensive discriminability are not salient to the representationist.

He argues that there are in fact differences in all of us concerning the symmetry of our colour spaces, especially between women and men, young and old, black and white, but these differences do not usually appear. This suggests a possible, alternative, kind of inverted spectrum argument that makes use of these fine-grained functional differences to justify actual, spectrum *shifts* that are prevalent between individuals in the real world. If the representationist does not have the means to deal with fine-grained functional differences, then this functional argument might be used against this representationist. This is precisely the line that Block (1999) takes in his latest argument:

> The representationism this paper is directed against is referential; the experience of red consists in its representing something as red. Although the kind of inverted spectrum needed to refute functionalism requires behavioural (and functional) isomorphism, representationism can perhaps be refuted empirically without these isomorphisms (p.41).

Block assigns to the representationist under consideration an appeal to a type of content that does not have the recourses available to handle the fine-grained functional differences under consideration. The contents he assigns to the representationist are limited to *expressible* concepts that cannot capture these fine-graineddistinctions:

> The representationist may suppose that the very evidence that I have appealed to for phenomenal difference also supports a representationaldifference. . . But it remains to be shown that we have visual representational resources capable of expressing such fine-grained differences. No doubt there are differences in dispositions which could be made explicit. . . But these representationalresources have

> been constructed, and the representationist has not shown that such resources are available at the moment of perception, i.e. on the fly (pp.45-46).

If Block's argument is sound, then the mere appeal to fine-grained functional differences is going to commit the representationist to an account whereby *inexpressible* contents are determiners for the type of content that is grounded by fine-grained functional differences. Therefore, if the representationist maintains that contents can be explicated by something like our course linguistic concepts, this representationism faces a problem when presented with Block's argument.

Asymmetry in the colour space is not the only possible fine-grained functional differences that might arise from inverted spectrum scenarios. Further subtle fine-grained functional differences might arise in the individual concerning brightness associations with specific colours, such as, light and dark (Tye, 1995b). At stage one of the inter-personal spectum inversion scenario, bananas looked bright, at stages two, three and four, bananas look dark. As noted earlier, the spectrum inversion of the type under consideration is a complete spectral inversion. However, if the inversion is merely a slight spectral shift, then these subtle associations present no problem to the above argument. However, if the spectrum is only shifted slightly, this increases the plausibility of the previous representationist reply based on phenomenal re-inversion in the individual after the lenses have been inserted.

The appeal to fine-grained functional differences as a response to inverted spectrums, is grounded on real experimentation, which is a plus in Gwendolen's eyes. If the representationist *only* appeals to fine-grained functional differences and if Block is right, then this rules out a type of representationism that holds that an individual's current linguistic concepts 'on the fly' are sufficient for determining the contents of their mental states. If the appeal to functional differences relies only on a break down of broader co-extensive discriminations, such as between purple, black and yellow, it is easy to modify the account so that it relies on a more moderate spectral shift rather than a complete spectral inversion; however, this does increase

the plausibility of spectral re-inversion as the individual interacts with his environment.

A representationist reply to inverted spectrums need not, however, appeal to phenomenal re-inversions or functional differences. A distinct representationist reply identifies phenomenal content or character with the narrow intentional contents of mental states; this allows for the possibility that an individual undergoing an a spectrum inversion has mental states with distinct narrow intentional contents whilst having mental states that have the same wide intentional contents (Tye, 1994).

1.3.4 Narrow contents

Tye in 1994 had a very different response to inverted spectrums, based on the idea of narrow content.[18] According to Tye the inverted spectrum argument does successfully refute wide intentional content representationism but does not refute narrow intentional content representationism. A colour inverted individual having, for example, a phenomenal content or character green while behaving in a 'normal red like manner' is not sufficient for this individual's states to represent the ripe tomato as red. The narrow intentional contents of that individual's perceptual experience represent the ripe tomato as green.

Tye justifies a narrow content, or individualistic, view of phenomenal content by formulating the following three Twin Earth scenarios: in the first scenario, Jones looks at a real tiger on Earth; in the second, Twin Jones on Twin Earth looks at a tiger that is genetically and biologically distinct from Earth tigers; on the third Twin Earth, an identical Jone's brain is stimulated

---

[18]This is in stark contrast to later Tye: "The lesson of the problem of transparency is that *phenomenology ain't in the head* . . To discover what it's like, you need to look outside the head to what brain states represent. Phenomenology is, in this way, externally based. So systems that are internally physically identical do not *have* to be phenomenally identical"(Tye, 1995b, p.151, italics his). "It seems to me, that we should resist the idea that phenomenal content is *narrow*, if by that is meant that phenomenal content is ultimately metaphysically fixed by what goes on physically inside the brain independently of everything else"(Tye, 1995b, p. 155, italics his).

so it is in the same psycho-chemical state as the other two. Tye subsequently argues:

> It follows from the argument I have given that the three experiences share a range of intentional features, namely those pertaining to directly observable properties (e.g., colour and shape). These features constitute the specific phenomenal character of the experiences. So, visual experiences have intentional features that are independent of the environments in which the subjects of the experiences are located. So, visual experiences have, in one clear sense of the term, narrow intentional contents (p. 168-9).
>
> But the narrow content of a perceptual experience does not exhaust its intentional content. For visual experiences. . . have wide contents too (p.169.).

This is demonstrated by considering Jones and twin Jones experiences of Tigers; thus:

> (They have) experiences with identical narrow contents (and thereby identical phenomenal characters) but different wide contents (in virtue of their different environments) (p. 169).

Then he concludes by explicitly rejecting Harman and Dennett's wide representationism.

> (This account) is in one respect, importantly different from the views of Dennett and Harman. For they maintain that the relevant content is wide. The result is that, while I reject visual Qualia, I still agree with such philosophers as Shoemaker and Block that what it is like supervenes on what is in the head (p.169).

Tye appeals only to an intuition about the supervenience of mental states on internal physical constitution to motivate his account of narrow intentional content. He applies his distinction to inverted spectrums thus:

> It may be objected that I haven't explained how Tom's experience can represent green when, as I noted above, it is an experience of the subjective sort that is normally produced in him by viewing red objects, and that normally produces in him the belief that something red is present. My reply should be obvious. wide functionalism cannot analyse the narrow content of visual experience, and this is a case of narrow content. So the tomato can look green to Tom *so long as his brain is not in the same physical state* (p.171, italics mine).

The success of this reply is dependent on the success of the notion of narrow content and whether it is applicable to phenomenal content or character. Tye's 1994 response to inverted spectrums does not sufficiently motivate a narrow content account based on his three Twin Earth scenarios. He offers a hastily argued account of narrow contents, grounded *only* on the intuition that narrow intentional contents supervene on intrinsic physical states. If individuals are spectrum inverted, they must be physically different, thus they have different narrow contents. But it is not clear that his appeal to his three Twin Earth scenarios is an adequate justification for narrow intentional *content.*

It is arguably as a response to Inverted Earth, a thought experiment considered in the next chapter, that a more plausible account of narrow intentional content might be thought to be applicable to phenomenal content or character. The kind of narrow intentional content account that will be presented is largely drawn from Jerry Fodor's (1990) account of narrow content, but instead of narrow intentional contents being functions that map contexts onto truth conditions, narrow contents will be defined as functions that map contexts onto contents. The narrow intentional content of a mental state is what is shared by the two mental states when they are transferredto distinct wide contexts. When two individuals are in the same context and they have mental states with the same narrow content, then they will have mental states that share a wide intentional content. In the case of the inverted spectrum scenario, the two individuals share a wide context, but it is understated as to whether they do, or do not, share narrow intentional contents; despite any physiological difference between the two individuals. If the two individuals, who, when in the same context share mental states with both narrow and wide intentional contents, are moved to different contexts, they will then have mental states that share a narrow content, but the wide intentional contents of their mental states will differ. This account of narrow intentional content will be developed and clarified in the next chapter as a representationist response to Inverted Earth.

1.3.5 Relativity

Another possible 'reply' to inverted spectrums is based on the idea that the phenomenal content or charactersof mental states might still be intentional content states, and they are relational, but not with respect to inputs and outputs. They might be relational with regards to each other, and this might be described as the Frege-Schlick view. "The Frege-Schlick view holds that qualitative properties are relational - part of a purely relational structure" (Stalnaker, 1999, p.392). Robert Stalnaker (1999) presents an analogy with examples of relational structures, one of these being a relational theory of space:

> There are, on this view, no absolute locations, one can still talk meaningfully about the spatial location of things, but this is just a framework for talking about the spatial relations between things. . . The analogies support, I think, the coherence of a purely functionalist account of qualia that takes phenomenal characterseriously, but treats it as a relational feature of our experience. Such a conception of qualia . . . might be grounded on . . . discriminatory capacities and judgements of similarity and difference (pp.393-4).

Stalnaker was *not* the first to apply such an analogy to qualia and inverted spectrums. As Austen Clark (1985) argues:

> Qualia identification is in many ways analogous to identification of places.[19] (Indeed, for the functionalist, qualia is a sort of spatial identification, where the "space" is that defined by the order yielded by qualitative similarities and dissimilarities).[20]. . . Corresponding to spectrum inversion is spatial inversion. One can imagine a universe which is spatially symmetric, so that relative to some locus every object has an exact spatial inverse. In such a universe, *two* places would satisfy the same spatial structural description, even when the spatial relations in which each stands are expanded to include every object in the universe. If the coordinate scheme is appropriately symmetric, every n-tuple of coordinates may be ambiguous in this way. But such a possibility fails to show that spatial location is not relational - that one can identify some "absolute" location independently of the network of spatial relationships in which it stands. . . If ambiguity persists, we can disambiguate by ostension. . .
>
> Similarly, even if the structure of qualitative similarities is symmetric, locations in it can be disambiguated by ostension to samples. . . Green qualitative contents are those qualitatively similar to those qualitatively similar to

---

[19]Clark attributes this analogy to Quine (1969, pp. 49-51).
[20]See Nelson Goodman (1977, pp 200-01).

> the one occasioned by that (pointing). This identification can succeed for everyone, even under the supposition of spectrum inversion. The friend of qualia may question whether such a description picks out the same qualitative content in everyone. Here too the response is that once such relations are fixed, there is no further fact of the matter; the respective sensations satisfy all the determinate senses of qualitative similarity and no other sense has been provided. To further query identity of qualia is analogous to querying identity of location, once spatial relations to some ostended place have been provided.
>
> Just as with spatial location, therefore, the functionalist holds to a relational theory of qualitative content. To specify the content of the sensory state is just to specify its place in a network of relationships of relative similarity and discriminability. To query its qualitative content in any more absolute way is as meaningless as querying location after co-ordinates have been given (pp.441-2).

Applying Stalnaker's analogy, it would be meaningless to compare one universe with another universe that was shifted three feet to the left. The relations are left intact, and therefore nothing really changes. To apply Clark's analogy, it would be meaningless to compare two locations between a spatially inverted universe with a universe the right way up. The relevance of these analogies should be fairly clear with regards to the inverted spectrum scenarios. Given two functionally identical individuals, there really are no grounds for claiming that their phenomenal characters are different if the relations between their mental states are kept intact. But what about the lenses in the intra-personal spectrum scenario? Surely these lenses do something to the phenomenal content or character of an individual.

However, according to the relativistic account, in the intra-personal scenario, the internal relations between an individual's mental states stay the same *despite the insertion of colour lenses.* These lenses maintain all the internal relations between mental states; so, for instance, when the individual at stage two has spectral inverting lenses inserted, all the phenomenal content or characterof his mental states only *appear to* be rearrangedwith respect to each other, but all the *relations* between the phenomenal content or charactersof his mental states stay the same; thus, the rearrangement of the phenomenal content or characters stays constant with regards to their

relations to each other. Therefore nothing phenomenal really changes after the immediate insertion of the lenses. There is no reason to believe that the individual will behave any differently from the way he did before these lenses were inserted. These lenses do not change the internal relations; therefore, the lenses change nothing.

A curious feature of this account is that the functions relied on are not functions from inputs to mental states to outputs, but functions that rely on some kind of internal similarity/dissimilarity relations, although to what degree these comparisons are internal is not clear. A phenomenal content is not a mental state that stands in a relation to an objective feature of the world, but 'something' that stands in a complex similarity/dissimilarity relation to every other phenomenal content. There is a sense in which phenomenal content or character is made redundant because there is no objective phenomenal content or character, only similarity/dissimilarity relations constitute the phenomenal content.

Is the above account of colour really plausible? The proponent of this relativistic account of colour appeals to his own brand of unsubstantiated metaphysical speculation. There is arguably a fairly strong intuition that colours really are monadic properties of objects or are constituted by monadic phenomenal contents ratherthan a complex relational property that any one colour really stands in a host of relations to other colours.

One reply to the above relational account is to use the asymmetry of colour spaces against such an argument and resurrect a spectrum shifted argument along the lines of Blocks. In other words allow that the lenses change the similarity and discriminability functions. This allows for a violation of coextensive discriminability. The discriminability functions are changed in a manner sufficient for there to be a change in phenomenal content or characterfor the individual after the lenses are inserted, but these changes are slight to not change any of the broader behaviours. Or if the cruder behaviours are changed, due to more substantial relational changes, after time the individual's crude behaviours revert to a pattern similar to before the lens insertion. This allows for the plausibility of the above relational analysis, but uses functionalism against the representationist.

1.3.6 Current wide-functional roles are not the only determiners of content

Another strategy for the representationist is to argue that *historical* wide-functional roles are also relevant. It is perhaps ironic that the representationists, who maintain that historic, wide-functional roles are determiners of content, are confronted by their own version of an inverted spectrum argument (Tye, 1995b, p. 206). It is possible that two beings could have perceptual contents determined by *distinct* historical functions; furthermore, these beings could in fact be physically identical.

> Now consider a creature that is a micro-physical duplicate of the first. The second creature is also a *narrow* functional duplicate. Might it undergo a different sensation?. . . , the answer to this is yes. All that we need to suppose is that the evolutionary histories and natural habitats of the two creatures are different, and that the brain states that realize sensations in the first creature are causally correlated with different external features from those the same brain states are causally correlated with in the second creature. In these circumstances, their sensations will have different representational contents and different phenomenal characters (Tye, 1995b, p.206).

Tye still insists that these creatures have to be in *distinct* 'natural habitats' so their mental states *also* have distinct current wide-functional roles.[21] Problems arise when historical wide-functional roles compete with current wide-functional roles. If phenomenal content or character is entirely determined by intentional content, which is only determined by historical long-arm functional roles, and if these historical long-arm functional roles are distinct in the two creatures, *and* if these beings are moved to the *same* 'natural habitat,' then these two creatures could still have distinct representational and phenomenal character despite having mental states with identical current wide-functional roles.

With only the relevance of wide-functional roles, it is not entirely clear how this point is applicable to the inter-personal spectrum inversion scenario. It

---

[21] I am assuming that when Tye refers to the causal correlations with external features, these occur both in the history of the creature and with current interactions of the creatures in their *distinct* habitats although this is ambiguous.

is hard to see what, if any, the historical long-arm functional roles of the mental states of an individual, plays, when he has had lenses inserted. This representationist position has, however, considerable relevance when the thought experiment of the next chapter is considered.

1.4 Conclusion

Just as Jack responds to his 'metaphysical speculation' about whether Gwendolen could only love a man named Ernest, 'metaphysical speculations' about inverted spectrums raise a whole host of possible types of representationists, as they attempt to respond to the argument. Perhaps the clearest type of representationist who faces a challenge from the inverted spectrum is the representationist who holds that current long-arm functional roles determine content. Is their theory shown to be false? No, but it looks as if, when confronted by the inter-personal spectrum inversion, they might well have to argue for some kind of phenomenal re-inversion. Alternatively they might try to appeal to fine-grained functional differences between both the individual at an early stage of his life and at the latter stage of his life. But these fine-grained distinctions must then be relevant to this person's wide intentional contents. If the representationist can only appeal to coarse functional differences that result in the application of broader linguistic concepts, 'on the fly,' then the appeal to fine-grained functional differences is not only redundant, it, ironically, suggests a way in which a *shifted* spectrum argument might be presented, not as mere metaphysical speculation, but as an argument against the representationist based on 'facts of life' (Block, 1999).

An alternative representationist response attempts to ground phenomenal character on internal similarity/dissimilarity relations. But these relations seem to ignore the importance of the relations between these states and the world; however, it is not easy for the phenomenist to adopt such a response to internal relations. But, it is not clear whether the representationist can consistently maintain such a relativistic account of colour given their insistence that colour really is a property of objects that constitutes the intentional content of the mental state. Another phenomenist response to these internal relations is to claim that the subtleties of these internal

relations may result in functional differences, but these fine-grained differences are not sufficient for establishing differences in intentional content. Finally the relativistic account relies on its own brand of pure 'metaphysical speculation'.

In the next chapter another 'metaphysical speculation' is considered: Inverted Earth. This metaphysical speculation motivates a representationism that appeals to narrow intentional contents as functions that map contexts onto contents. Inverted Earth also motivates a representationist account of intentional content that relies on historical long-arm functional roles. Ultimately the Inverted Earth argument will be demonstrated to fail as a conclusive refutation of representationism, but Inverted Earth is instrumental in showing how some of the representationist replies to inverted spectrums, are not so easily applicable to Inverted Earth. Indeed some of these replies to inverted spectrums motivate the very premises of the Inverted Earth argument against the representationist.

Chapter 2: Inverted Earth (I)

2.1 Introduction

In the last chapter various replies to inverted spectrum scenarios were considered and distinct types of representationism were introduced emerging out of possible replies to the inverted spectrum scenario. There were some representationist replies that by their *non*-engagement with inverted spectrums left much to the imagination as to the type of representationism being sustained by the objectors; however, there were some types of representationism that were more revealing. A 'linguistic-conceptual' type of representationism emerges out of a reply to inverted spectrums based on the meaninglessness of sensation terms *if* sensation terms are taken to only refer to some kind of private inner object. It is precisely this type of consideration that partly motivates Block's presentation of the Inverted Earth argument (Block, 1990). A current long-arm functional role representationism also motivates Block's presentation of inverted spectrum arguments and his presentation of Inverted Earth; that is, current long-arm (or wide) functions of mental states determine the intentional contents of mental states.

In this and the next chapter, the representationist replies to inverted spectrums will be critically evaluated with specific reference to their replies to Inverted Earth; thus elaborating on some of the representationisms already encountered. For example, is an appeal to phenomenal re-inversion applicable to Inverted Earth? Two further types of representationism, partly motivated as possible replies to *Inverted Earth*, will be considered: narrow content representationism and *historical* wide functionalist representationism. The former will be discussed in this chapter, the latter in the following chapter. The narrow content representationist account of phenomenal characterwill be demonstrated to fail on grounds independent of Inverted Earth. The historical wide representationist response will be rejected for the following reasons: first, because of problems inherent with these types of accounts when applied to phenomenal content; and second, because of the availability of a superior response to Inverted Earth. Finally, in the next chapter, this superior wide representationist response to Inverted

Earth will be presented as the most adequate representationist response to Inverted Earth.

2.2.1 Inverted Earth

Inverted spectrums were presented in the last chapter as an attempt to describe a case where an individual might have identical intentional contents to his mental states as another individual/or himself at an earlier stage, but these two individuals have distinct phenomenal contents or characters to their mental states. Block (1990, 1996) presents Inverted Earth as an attempt to describe an intuitively plausible scenario whereby an individual has distinct intentional contents with another person, or himself at an earlier stage, but has identical phenomenal content to his mental states.

> I will describe a case . . . of inverted intentional content . . . combined with identical qualitative content. . . I will describe a case of two persons/stages whose experiences are qualitatively the same but intentionally . . . inverted (Block, 1990, p.62).

For example, an individual, Ed, shares the intentional content of his mental states with *another* individual, Ted. Ted, however, clearly has mental states with intentional contents distinct from the intentional contents of Ed's mental states *at an earlier time in Ed's life.* Thus, the intentional content of Ed's mental states change between an early stage in his life and a later stage in his life. However, the phenomenal contents or character of Ed's mental states stays the same throughout his life. Therefore, intentional content is distinct from phenomenal content or character; representationism is false.

The phenomenal contents or character of Ed's mental states stays the same for two reasons. First, Inverted Earth is a place where all the colours of objects *are* inverted: Bananas really *are* purple/blue and fire engines *are* green, grass *is* red and the sky *is* yellow. Second, Ed travels to Inverted Earth, but has lenses inserted that *re-invert* the look of all the colours of objects.[22] Therefore, the phenomenal character of Ed's mental states stays

---

[22] All Ed's colour pigments are changed so that his skin does not look a funny colour, as are his bodily fluids.

the same. Nothing changes phenomenally for Ed after he arrives on Inverted Earth. For Ed the sky looks just as blue and the grass looks just as green. Block (1990) describes the inversion process and the constancy in phenomenal content or characterin the Earthling's mental states thus:

> A team of mad scientists knock you out. While you are out cold, they insert color inverting lenses into your eyes, and change your body pigments so you don't have a nasty shock when you wake up and look at your feet. They transport you to Inverted Earth. . . You wake up, and since the inverting lenses cancel out the inverted colors, you notice no difference at all. . . For example the Yellow sky looks Blue to you. . . As far as the qualitative aspect of your mental life is concerned, nothing is any different from the way it would have been had you stayed at home (p.63).

Why does the intentional content of Ed's mental states change? The type of representationist under consideration determines the specific motivation for thinking that the intentional content of Ed's mental states change.

2.2.2 Linguistic agreement

Let us start with a type of linguistic-conceptual or 'semantic' representationism motivated out of the 'anti-private language' reply to inverted spectrums. Ted is an inhabitant of Inverted Earth; however, he is part of a linguistic community that has an unusual language. Ted's colour terms are inverted relative to Earth: what we would call 'red,' Ted would call 'green,' what we would call 'yellow,' Ted would call 'blue'. Ripe tomatoes on Inverted Earth do happen to be green; therefore, Ted describes ripe tomatoes on Inverted Earth as 'red.' The curious implication of all this is that when Ed arrives on Inverted Earth and has lenses placed in his eyes, he appears to be in complete linguistic agreement with Ted concerning his usage of all his colour terms and the way these colour terms are ascribed to objects. Why does linguistic agreement matter to the content of mental states? Well, if we recall from the last chapter, it might be argued that Wittgensteinian considerations suggest that colour terms do not refer to 'inner private objects,' but only derive their meaning from use within a linguistic community in a shared environment. Ed and Ted share a language that agrees with the meaning of their colour terms and they share an environment. However, comparing Ted to Earthling's on Earth, there

appears to be a substantial linguistic disagreement between Earthlings and Invertlings as to the meaning of colour terms. Therefore, the referents of Ed's colour terms must change.

If this is so, and the 'semantic' representationism is assumed, then the contents of Ed's mental states necessarily change. His belief that ripe tomatoes are 'red' must change since he now belongs to a linguistic community where the meaning of the term 'red' has changed. Maybe it takes a while to become embedded in a linguistic community before there is complete linguistic agreement. However, if so, this is no threat to the argument. This is no threat to the argument because all that has to be done to modify the scenario is to allow Ed to stay on Inverted Earth until he becomes embedded in his new linguistic community.

> Nonetheless, according to me, after enough time has passed on Inverted Earth, your embedding in the physical and linguistic environment of Inverted Earth would dominate, and so your intentional contents would shift to be the same as those of the natives (Block, 1990, p.64).

But, however long Ed stays, the phenomenal character or content of his mental states will not change; whilst the intentional contents of his mental states will change.

So, one major motivation for thinking that Ed *shares* intentional contents with Ted is grounded on the anti-private language considerations to which the representationist might have appealed in order to rule out inverted spectrums. The *difference* between the intentional content of Ted's mental states and Ed's mental states at an earlier stage in Ed's life are also grounded on the anti-private language considerations that the representationist might have appealed to in order to rule out inverted spectrums. Therefore, a *change* in the intentional contents of Ed's mental states is also motivated out of anti-private language considerations. However, there are extremely good grounds for thinking that the phenomenal character or content of Ed's mental states *do not change,* between himself at an early stage of his life and himself at a later stage in his life, because of the lenses. Thus, the intentional content of this individual's mental states change whilst the phenomenal characteror content

of this individual's mental states do not. Therefore, intentional content does not entirely constitute phenomenal content or character; 'semantic' representationismis false.

2.2.3 Current long-arm functional roles

However, anti-private language considerations are not the only motivation for arguing that the intentional content of Ed's mental states change. If the representationist maintains that it is the current long-arm functional roles that determine content, then the Inverted Earth scenario *need* make no appeal to either anti-private language considerations or a third party individual in order to establish a change in intentional content in Ed's mental states.

Ed at an early stage in his life has mental states with certain current functional roles and Ed at a later stage in his life has mental states with distinct current functional roles. An appeal to the current wide functional roles of Ed's mental states thus motivates a change in content in Ed's mental states. On Earth, the mental states of Ed, which once tracked ripe tomatoes, fulfilled a certain current function that enabled him to describe red tomatoes as 'red' and identify red objects. The inputs to Ed's mental states with red intentional content, on Earth, were red objects such as ripe earth tomatoes, but after the move to Inverted Earth, the current mental states that once served to identify ripe tomatoes as 'red', now function to identify green objects. The inputs to Ed's mental states of ripe tomatoes on Inverted Earth arise from green objects; thus the inputs to Ed's mental states change. The outputs of Ed's mental states also change because the mental states that once caused an interaction with red objects now cause interactions with green objects. Both the inputs and the outputs to Ed's mental states change after the move to Inverted Earth; thus the current long-arm functional roles of Ed's mental states change after he moves to Inverted Earth:

> [Y]ou are in a state produced in a natural and immediate way by blue things, one that plays the aforementioned familiar role controlling your responses to blue things. At the same time, Twin [or yourself at an earlier time] is in a state produced by yellow things, one that plays the same familiar functional role in controlling his [or yourself at an earlier time] responses to yellow things (Block, 1990, p.66).

The current long-arm functional roles of Ed's mental states change; thus the intentional content of Ed's mental states changes. *Yet*, the phenomenal content or character of Ed's mental states does not change because of the inverting lenses; therefore, current long-arm functional representationism is false.

Alternatively, a third party might be introduced to demonstrate a change of intentional content based on long-arm functional roles. Ted might be re-introduced. Ted, the Inverted Earthling, has mental states with a current functional role identical to Ed's after Ed arrives on Inverted Earth. This is because the kinds of input that enters Ted's mental states when Ted looks at a ripe tomato on Inverted Earth are the same kinds of input that enters Ed's mental states when Ed looks at a ripe tomato on Inverted Earth. The output of Ted's mental states when he responds to the ripe tomato are the same as the outputs of Ed's mental states when he responds to a ripe tomato on Inverted Earth. There is no functional difference between Ted and Ed. However, Ted's mental states serve a current functional role clearly distinct from the functional role Ed's mental states *used to play*. Therefore, the current long-arm functional role of Ed's mental states changes; thus, according to a representationist that holds that current long-arm functional roles determine intentional content, Ed's intentional contents change. *Yet*, Ed undergoes no phenomenal change because of the lenses. Therefore, intentional content does not entirely constitute phenomenal content or character; current wide representationism is false.

Just as one might consider linguistic embedding to take some time, after Ed arrives on Inverted Earth, one might argue that a change in the long-arm functional roles of Ed's mental states takes some time to establish after Ed arrives on Inverted Earth. It might be argued that it is not the immediate inputs and outputs of mental states that determines the intentional content of these mental states, but causal co-variations between the mental states and what they 'track' now *and what they used to track* determines the intentional content of mental states. These causal co-variations might take

some time to establish.[23] This kind of account of intentional content begins to move away from what might be described as *current* long-arm functional representationism. It is how mental states have functioned in the past that begins to be relevant as to the intentional content of these mental states. Once again, however, this is no threat to the Inverted Earth argument. It merely has to be maintained that however long Ed stays on Inverted Earth, even after his mental states have started to co-vary with objects on Inverted Earth, the phenomenal character of Ed's mental states stays constant.

2.2.4 Summing up the Inverted Earth argument

It appears difficult for the representationist to argue that the phenomenal characterof the Earthlings mental states does not stay the same after the move to Inverted Earth because of the strange colours on Inverted Earth, and the insertion of colour-inverting lenses into the eyes of the Earthling. Therefore the first premise is established: *there is no change in the phenomenal content or character of the Earthling's mental states between before he arrived on InvertedEarth, after he arrived on Inverted Earth and for however long the Earthling stays on Inverted Earth.* However, the intentional content of the Earthling's mental states changes.

The intentional content of the Earthling's mental states changes because he becomes embedded in a new linguistic community on Inverted earth where colour terms refer to different colours. If there are no 'private inner objects' or intrinsic properties to which these colour terms refer, then it must be assumed that as the Earthling's colour terms start to become synonymous with the language of his current community, the meanings of his colour terms change. The intentional contents of his mental states likewise change. Alternatively the intentional content of the Earthling's mental states changebecause the current long-arm (or wide) functional roles of his mental states change. If current long-arm functional roles determine intentional content, then the intentional content of the Earthling changes. The second premise is established: either by the 'anti-privatelanguage' reply

---

[23] See Dretske (1981,1995) for an account of causal co-variation determining intentional content.

to inverted spectrums, or by the representationist appeal to the current long-arm functional roles of mental states; *the intentional content of the Earthling changes either after his immediate arrival on Inverted Earth, or however long it takes to become embedded in the linguistic community, or however long it takes for his mental states to start to causally co-vary with the properties of objects in his new environment.*

The phenomenal content or character of the Earthling's mental states stays constant; the intentional content of the Earhtling's mental states changes; therefore, representationism is false.

2.3.1 The representationist replies

There appears to be two general strategies open to the representationist: first, to deny that the intentional content of the Earthling's mental states really changes. If the intentional content does not change then, there is no problem with the phenomenal character of the mental states staying constant. Perhaps there is something in the *history* of the Earthling that fixes the intentional content of the Earthling's mental states. Alternatively, perhaps the Earthling's *narrow* intentional content stays the same while the Earthling's *wide* intentional content changes. These replies will be discussed in due course.

The second representationist strategy is to deny that the phenomenal content or character of the Earthling stays constant after he arrives on Inverted Earth. If the phenomenal content or character of the Earthling's mental states re-inverts, as the intentional content of the Earthling's mental states changes, then there is no reason to assume that phenomenal character is anything over and above intentional content. If the representationist maintains that the intentional content of the Earthling's mental states changes as soon as the earthling arrives on Inverted Earth, then there must be an immediate phenomenal re-inversion. If the representationist maintains that the intentional content of the Earthling's mental states slowly changes after he lives on Inverted Earth for a while, then there must be a slow phenomenal re-inversion.

2.3.2 Phenomenal re-inversion and 'semantic' representationism

The 'linguistic-conceptual' or 'semantic' representationist might employ phenomenal re-inversion to correspond with the changes in the meanings of the Earthling's concepts. This, however, looks somewhat implausible because it is not clear as to what the Earthling's colour terms refer to, as the Earthling's intentional content proceeds around the spectrum to get back to normal. If 'something' is creeping back to normal, then what is this 'something' that the colour terms of the Earthling are supposed to be referring to? To employ the semantic representationist's refrain: "this something is a nothing."[24] But if there is a nothing, then nothing slowly creeps around the spectrum. If nothing slowly creeps around the spectrum, then there is no phenomenal re-inversion. Slow phenomenal re-inversion is just not consistent with this type of representationism.

If there is no slow adapting 'around' the spectrum, then it looks like this kind of representationism requires an immediate and complete switch in intentional content as soon as the meaning of the Earthling's colour terms changes. But is an immediate switch really plausible only because an individual slowly becomes embedded in a distinct linguistic community? Arguably not.

> However, it seems extremely implausible that Ned's *experience* of these colours would undergo a corresponding shift: his qualitative experience doesn't flicker, or become indeterminate, or finally reverse, as the semantics of his words and thoughts might well do, simply as a result of conversing with Inverted Earthlings (Rey, 1998, p.443).

To put further strain on the representationist who appeals to anti-private language considerations, imagine if Ed was to return to *Earth* with an Invertling, Ted. Ted would disagree with the native Earthlings, describing a banana as 'blue'; Ed would agree with Ted that the banana is 'blue'. This would be so if Ed's perceptual processing mechanisms have somehow counteracted the effect of the lenses; thus he would see it as yellow. He has fully adapted to the Invertling's language, so he calls yellow things 'blue'.

---

[24] They might claim that it is a 'something', but according to the semantic representationist it is certainly not an inner object of awareness.

If he has not 'adapted' to his lenses and phenomenally re-inverted, he would also describe the banana as 'blue'. He would see the banana as blue and describe it as 'blue'. Now imagine that a native Earthling, Fred, puts on colour inverting lenses. If Fred talks to Ted his colour terms means one thing, if he talks to a native Earthling his colour terms mean another thing, but *it is simply not clear how being part of one linguistic community as opposed to another linguistic community has any effect on the phenomenal content or character of the mental states of Fred; it is the lenses that matter.*

This representationist might attempt to appeal to fine-grained functional differences between the Earthling and the Invertlings to argue that they never really mean the same things despite the apparent agreement in their use of colour terms. But it is not clear as to how fine-grained functional differences in the use of colour terms is salient to the meanings of colour terms. Therefore, I am going to agree with Georges Rey and also argue that this type of representationism does not have the resources to handle Inverted Earth because: "phenomenal similarities and differences don't track wide semantic similarities and differences" (Rey, 1998, p.442). However, other types of representationism might also appeal to phenomenal re-inversion.

2.3.3.1 Phenomenal re-inversion and current long-arm functional representationism

The current long-arm functional representationist might also appeal to phenomenal re-inversion, just as they appealed to phenomenal re-inversion in the intra-personal spectrum inversion scenario discussed in the last chapter. First Michael Tye (1998):

> Reflecting further, am I going to insist that there is no phenomenal difference between my present visual experience and my earlier one? . . . Indeed, if the 'I' of the thought experiment is myself, then the answer is 'Certainly not.' Since, in my view, the phenomenal character of any phenomenal state . . . is a matter of representational content, it immediately follows that there is a difference between the phenomenal character that is presently accessible to me, as it were, and the original phenomenal character of my visual experience at t (given the representational change) (pp.470-1).

- then Brendon Lalor (1999):

> The interactivist response to the Inverted Earth argument, then, is clear: deny Block's intuition that ripe tomatoes would look red to his twin; they would come to look green, since that's the color of tomatoes on Inverted Earth. Their looking green to him is just a matter of the information-bearing states in him being linked to the world so as to signal green under what have become normal conditions (p.269).

It is possible to delineate different types of phenomenal re-inversion. The first type of phenomenal re-inversion is either an immediate total phenomenal re-inversion as soon as the Earthling opens his eyes, for the first time, on Inverted Earth, or an immediate and total switch that occurs at some later time after the Earthling's arrival on Inverted Earth. Either he really sees the colours of objects for what they really are, as soon as he arrives on Inverted Earth, or a total instantaneous phenomenal re-inversion occurs at some point later on in his stay on Inverted Earth; one moment he sees things as they looked as on Earth, then the next minute he sees the colours as they really are. The second type of re-inversion is a slow shifting around the spectrum as the look of objects creeps back, to reflect the actual colours of objects.

2.3.3.2 Immediate and total phenomenal re-inversion

The problem with the immediate and total phenomenal re-inversion reply, when applied to Inverted Earth, is that as soon the Earthling arrives on Inverted Earth and opens his eyes, he has had lenses inserted into his eyes! The whole point of the lenses is to change the look of the objects so that objects appear the way things once used to look to the Earthling back on Earth. According to this representationist reply, the very existence of the lenses begs the question against the Represenationist, but what is so troubling about these lenses? It is worth recalling the relativistic account of colour discussed in the last chapter.

If colours are only determined by their relations of similarity and differences to other colours, then this has some curious implications for Block's attempted construction of the Inverted Earth scenario. First, it would not be

possible to even paint objects *different* colours and keep their relations to one another intact. Second, it would not be possible to create lenses that would *change* the look of colours and keep the colours' relations to one another intact. This is the case not only because our colour spaces are asymmetrical, although this does present a problem of keeping fine-grained functional differences intact, but also because, according to the relational theory, being a colour just *is* its relations of similarity and differences to all other colours. Block's scenario is, therefore, doubly impossible according to a relational theory of colour for the following reasons: first, because of the impossibility of painting objects different colours and keeping their relations intact; and second, because of the impossibility of creating lenses that invert the looks of colours whilst maintaining their relative relations of differencesand similarities. Not only do the colours of objects on Inverted Earth *not* change relative to the colours of objects on Earth, the lenses do nothing. If the relativistic account of colour is accurate, then the inverted Earth scenario is equivalent to a complete and total phenomenal re-inversion as soon as the Earthling arrives on Inverted Earth. Seeing as there is no phenomenal inversion in the first place, the colours of things on Inverted Earth look no different from the colours of things on Earth but not because of the lenses.

However, one obvious reply to this reply is to doubt the plausibility of a relational theory of colour. If the relational theory is assumed, it should be possible to start with a colour wheel and change one colour segment of the wheel so that it looks different, then every other segment could be changed, one at a time, so that all the original relations are kept intact. But by the time the painting of the colour wheel is finished, the change in colour of the original segment would no longer look any different from the way it looked before it was painted. This does not look intuitively plausible.

Another alternative for the proponent of the Inverted Earth argument is to bow to the pressures of a relational theory of colour and allow that painting objects different colours will not maintain relations of similarity and differences. The Earthling does not undergo a spectrum inversion, but a spectrum shifting. The lenses will somehow *restore* these relations so that objects will continue to look, to the Earthling, just the way they looked on

Earth. If the Earthling were to remove his lenses, the colours of objects would look very different albeit not symmetrically inverted. We must remember that all that is required is that the phenomenal content or character of the individual's mental states stays the same while the difference in the colours of objects is sufficient to establish a difference in intentional content of this individual's mental states. There will, however, remain fine-grained functional differences between the Earthling's and the Invertlings.

Finally, it is perhaps not clear that the representationist under consideration can consistently maintain that colours are *not* objective monadic properties of objects. If it is accepted that colours are objective monadic properties of objects that are not determined relationally, with regards to one another, then the lenses are a realistic possibility; thus the plausibility of immediate and total phenomenal re-inversion is highly suspect.

2.3.3.3 Total and complete phenomenal re-inversion or a slow re-inversion some time *after* the Earthling arrives on Inverted Earth

If the representationist maintains that it takes a while for the mental states of the Earthling to functionally establish themselves in their new environment, the total and instantaneous phenomenal re-inversion or the completion of the slow phenomenal re-inversion might occur some time after the arrival of the Earthling on Inverted Earth. It is only when the phenomenal re-inversion occurs that the intentional contents of the Earthling's mental states change.

> [T]hey would *come to* look green, since that's the color of tomatoes on Inverted Earth (Lalor, 1999, p.268, italics mine).

Brendon Lalor appeals to the slow variety of phenomenal re-inversion. However, the question remains as to what exactly is the specific cause of the phenomenal re-inversion? He supports his claim partly with an appeal to the plasticity of our visual systems and more specifically a gesture towards experiments involving *spatially* inverting goggles:

> [I]n the much-discussed studies on visual field inverting goggles, the world comes to seem normal to many visually inverted subjects to the point that they can go

> about their lives as usual. Subjects who adjusted most thoroughly did not think of their visual fields as upside down. In all of these cases, after recalibration, informational properties are restored or preserved, and with them, experiential properties (p.269).

According to Lalor's interpretation of the data, concerning spatial goggle inversion, the wearers of these goggles, over time, phenomenally re-invert the spatial inversion.[25] If spatial re-inversions can occur, why could not the Earthling's mental states compensate for the colour inverting lenses, thus phenomenally re-inverting in a way that corresponds to the change in intentional content of the mental states? According to this reply, given the plasticity of our visual system, all of which is supported by impressive empirical research(Gwendolen would be pleased), it is perfectly possible to endorse some kind of slow phenomenal re-inversion in the Earthling. Indeed we might be compelled to accept some kind of phenomenal re-inversion in the Earthling.

However, the Inverted Earth scenario is set up in such a way that the plasticity of the visual system of the Earthling is simply not relevant. In the Inverted Earth scenario, there is no opportunity for the plasticity of the Earthling's visual system to assert itself. If we consider the spatially inverting goggles, the individual wearing the goggles has to substantially alter his basic motor actions in order to perform even simple tasks; for example, catching a ball or making a cup of tea. Furthermore, it is arguably the agent's self motivated behavioural gross re-adjustment, in order to perform simple tasks, that is required for there to be any re-inversion. (Held & Blossom, 1961).[26] But the Earthling who wears the colour inverting

---

[25] However, this is by no means clearly established by the existing experimental data on inverting lenses (see Dolezal, 1982, pp.227-8). But for the sake of argument let us assume that Lalor's interpretation is accurate.

[26] "Previous experiments have shown that errors in hand-eye coordination induced by this prism rearrangement were partially compensated only after S had moved his hand under the rearranged condition. A more interesting result . . . was the finding that after S's passive hand and arm had been moved in a comparable manner by E, no compensation occurred. . . According to these results, not simply movement but self-produced movement with its contingent reafferent stimulation is the critical factor in compensation for rearrangement. . . Experimental procedures showed that full and exact compensation for these errors requires gross bodily movement and more specifically, self-produced movement for prolonged periods of exposure " (Held & Blossom, 1961, pp. 34, 37).

lenses *requires no self-motivated behavioural re-adjustment in order to cope with his new environment on Inverted Earth.*

For the Earthling, everything looks just the same way as it did on Earth. Making a cup of tea on Inverted Earth is just as simple as making one on Earth. Even if the Earthling were to become an art historian, he would have no problems describing the colours of paintings in a way that would not raise an eyebrow with any of his colleagues. Notice that in the case of the intra-personal spectrum inversion, the individual, after the lenses were inserted, might have to behaviourally re-adjust to fit into his linguistic community; therefore, Lalor's reply based on the plasticity of our visual systems *might* be applicable to inverted spectrum scenarios,[27] but it is certainly *not* applicable to Inverted Earth.

So a deeper question remains about the Earthling on Inverted Earth: What possible *change* in his mental states motivates a possible change in the phenomenal content or characterof this individual's mental states? He does not have to change his language. No self-motivated behavioural re-adjustment is required to perform even the simplest of tasks. Presumably, from the retina inwards there is absolutely no change in the internal processing of the Earthling's brain states. Plausibility suggests that the phenomenal content or character of this individual stays the same. However, the wide representationist consistently claims that the only change is in the relations of the individual to his environment and none of the internal states need change, yet according to the wide representationist the relations matter.

Curiously, however, a further implication of the wide representationist reply, which relies on phenomenal re-inversion, is that because the intentional content of the Earthling's mental states is constituted by only the long-arm functional roles or the Earthling's mental states, the internal functional roles of the Earthling's mental states necessarily do not change. If phenomenal re-inversion occurs, the Earthling *must not notice.*

---

[27] Although it is not clear what gross behavioural re-adjustments are required to phenomenally re-invert if ones *colours* are inverted; perhaps if one worked in a paint factory moving pots of paint around and matching them to labels.

Assuming that the wide representationist still maintains that there is a phenomenal re-inversion without any appeal to empirical evidence, but only because phenomenal re-inversion is an empirical *possibility*, they still have to account for the phenomenal characterof the Earthling changing *without the Earthling noticing.*

2.3.3.4 Unnoticed phenomenal shifts

Given the supposition that the Earthling fails to notice the phenomenal shift, and indeed this is required if the representationist maintains that is only the wide functions of the mental states that count when it comes to intentional contents, there is arguably a substantial *prima facie* burden on the representationist to explain how this phenomenal re-inversion could occur without the Earthling noticing. This requirement most likely will rule out a complete spontaneous re-inversion at some point in the Earthling's stay on Inverted Earth. Slow re-inversions that creep around the spectrum look like the most plausible option for the representationist. A failure in short term memories of colours would also be helpful to the representationist.

> The subject never notices a change in qualia because of slow compensatory adjustments to the mechanism which recall memories of how things used to look (Lalor, 1999 p.271).

Arguably, the reason why the shifting has to be slow is that this might allow for the introduction of an argument based on the well-established fact that we often have trouble distinguishing between distinct but close shades of colour (Lycan, 1996).

2.3.3.5 Transitivity failure

If the proponent of the Inverted Earth argument argues that the phenomenal content or characterof the Earthling's experience of, for example, the sky does not change after the Earthling arrives on Inverted Earth and stays on Inverted Earth for a considerable period of time *only* because the sky's shade of blue looks just like the shade of blue it looked like a moment/week/year ago, then William Lycan (1996) argues that the supporter of the Inverted

Earth argument relies on a false transitivity principle to establish this phenomenal constancy. Transitivity failure is sometimes shown to fail across the distinguishability of close colours on a patch, in this case, the colours are not presented adjacent to each other on a patch, but temporarily, one after the other.

> Block needs a longish interval during which all the relevant representational content change but colour sensations remain introspectively indistinguishable. Perhaps the representationalcontents do gradually change, but what shows that the colour sensations' Strange Qualia . . . do not change with the representational contents? I suspect that a transitivity principle is at work here: if e1 is introspectively indistinguishable from e2 and e2 is introspectively indistinguishable from e3, then e1 is introspectively indistinguishable from e3, and introspective indistinguishability is maintained over a long interval by way of very short experience-pair intervals (Lycan, 1996, p.128).

Lycan claims that the proponent of the Inverted Earth argument is committed to arguing that the phenomenal characterof the Earthling stays the same after his arrival on Inverted Earth *only* because of the following argument: The sky looks blue on the first day the Earthling arrives on Inverted Earth. On the second day, the sky looks just as blue to the Earthling as it did on the first day. On the third day, the sky looks just as blue on the second day. Therefore, by the *phenomenist* applying a false principle of transitivity, the *phenomenist* infers that the sky on the third day looks exactly the same shade of blue to the Earthling as it did on the first day. According to Lycan the *phenomenist* assumes transitivity and re-iterates it for as long as they care; thus establishing that the phenomenal characterof the Earthling's mental states *do not* phenomenally re-invert.

After Lycan commits the anti-representationist to this transitivity principle, he rejects the transitivity principle; thus, according to Lycan, re-opening up the plausibility of phenomenal re-inversion without the Earthling noticing:

> I do not grant such a transitivity principle . . . notoriously there can be a just noticeable difference between a and c even though there is no just noticeable difference between a and b or between b and c (p.128).

According to Lycan the *only* motivation for maintaining a constancy in phenomenal characteris undermined; therefore, there are no grounds for the anti-representationistto maintain that the phenomenal content or characterof the Earthling's mental states stay the same while the intentional contents of the Earthling's mental states differ.

However, it is not clear that the phenomenist needs to deny Lycan's transitivity principle in order to justify why there cannot be unnoticed phenomenal shifts in the case of Inverted Earth. Lycan mistakenly assumes that the phenomenist relies on a false transitivity principle for discrimination of close shades to establish that there is no change in the phenomenal content or characterof the Earthling. Lycan, however, ignores that failure of transitivity is applicable only to *close* shades. It is surely plausible that transitivity *holds* for widely distinct shades of colour. Yellowish orange is clearly distinguishable from reddish orange; reddish orange is clearly distinguishable from red; thus it is also fairly plausible that yellowish orange is clearly distinguishable from red. Furthermore, anyone who argues for an unnoticed phenomenal re-inversion is not just committed to a failure of transitivity of close shades but a re-iteration of this failure to get from one shade to a very different shade; and this is arguably absurd as David Chalmers (1996) points out:

> It is true that there can be unnoticeable differences between different experiences. If one changes a shade of red little enough, I will not be able to tell the difference. . . but if this were all that was going on one could re-iterate such a change a thousand times, eventually showing that red and blue produce the same experiences, which is ridiculous. So there can be *some* difference in experience that is not noticeable. One can observe this phenomenon by looking at a wide expanse of paint of subtly varying shade; sometimes it is extremely difficult to tell whether one's experienceof different parts is the same or different. But importantly unnoticeable differences are very *small* (p.267).

It is precisely the large kinds of colour shifts that Lycan thinks can un-problematically occur without the agent noticing through a series of re-iterated failures of transitivity. Lycan simply assumes that a denial of transitivity of close shades is sufficient for claiming that an individual when perceiving a shift from orange to red will not at some point notice that there

has been a change in his phenomenal content or character; surely this is mistaken.

Furthermore, transitivity is irrelevant as a justification for a constancy in the phenomenal character of the Earthling's mental states because transitivity does not need to be invoked in order to argue that there is no shift in the phenomenal content or character of the Earthling's mental states. The phenomenist has independent grounds for maintaining that there is no un-noticed phenomenal re-inversion. These will be discussed in due course. The representationist needs to offer a far more convincing argument to justify a slow unnoticed phenomenal re-inversion.

2.3.3.6 Memory failure

The representationist might argue that the only ground for the proponent of the Inverted Earth argument to hold is that the phenomenal character of the Earthling stays constant because of the reliability of the Earthling's memory. The Earthling remembers that the sky looks just as blue as it once did, therefore the phenomenal character of the Earthling does not change. But why should we take the memories of the Earthling at face value? Indeed even according to the proponent of the Inverted Earth argument there must be something going wrong with the Earthling's memories, if the intentional contents of the Earthling's memories do change. First Lycan (1996):

> On Block's own view, memory contents will undergo the representational shift. So when you *say* or think to yourself, "Yes, the sky is as blue as it was thirty years ago," you are not expressing the same memory content as you would have when you had just arrived on Inverted Earth. You are now remembering or "remembering" that the sky looked yellow, since for you "blue" now means yellow. And that memory is *false,* since on the long-ago occasion the sky looked blue to you not yellow (p.130).

- then Tye (1998b)[28]

> My report of no change in phenomenal character. . . is necessarily in error. . . When I now say after a long stay on Inverted Earth, "grass looks just as green to me now,

[28] This is hinted at in Tye (1995b).

> just as it did five, ten, and twenty years ago," I am wrong. 'Green' (in Inverted English) means red; and grass did not look red to me twenty years ago. My memory has led me astray (p.466).

- then Lalor (1999):

> What about the memory of your Earth sky and the look of the sky you enjoy as an assimilated Inverted Earther? Block thinks these will be qualitatively the same; but there is no reason to suppose this if intentional adjustment is complete. It is true that you cannot tell the difference between the phenomenal content of your memory of Earth's blue sky and your current perception of Inverted Earth's yellow sky. In one sense, your memory represents a blue sky; that's what was encoded. But in another sense it does not represent the sky as blue since, as you now decode the 'blue' memory, you can't help but unzip it as yellow, and thus misremember. Given the connection between color perception and memory, the adaption of your perceptual system has limited your ability to correctly recall and elaborate the remembered color of objects. Yours is a memory of a blue sky which represents that sky as being yellow (p.280).

Lycan, Tye and Lalor attribute to the Earthling a major breakdown in the Earthling's phenomenal memories. This is because when the Earthling claims that the sky looks just as 'blue' as it once did, the Earthling means that the sky once looked yellow. But, the sky never looked yellow. Therefore, there is a major breakdown in the Earthling's phenomenal memory. If there is such a substantial breakdown in the memories of the Earthling, there are no real grounds for appealing to the reliability of the Earthling's phenomenal memories in order to establish that there is no phenomenal change in the Earthling's mental states after a long stay on Inverted Earth. Phenomenal re-inversion remains a viable reply to Inverted Earth.

However, it looks as if the proponent of the Inverted Earth scenario has some plausible replies to the above claims of memory failure. First, it appears gratuitously *ad hoc* to just claim that there is a break down in one aspect of an individual's memory; therefore, one cannot rely on any of the Earthling's memories of how objects used to look to him. Second, if the Earthling is aware of the fact that lenses has been inserted into his eyes, and it is assumed that the individual does not undergo any phenomenal re-

inversions, then it is perfectly possible to make sense of his memories.[29] Allowing for this modification, the Earthling accurately remembers that the *colour* of the sky looks the same way as it looked when he first arrived on Inverted Earth and the same way as a different sky looked to him when he was on Earth. The Earthling can consistently maintain that the sky on Earth never looked 'blue,' meaning yellow. It once looked 'yellow' meaning blue, but it did look the same *colour*. We can therefore disambiguate the following distinct memories that the above representationist claims fail to disambiguate.

a) The *colour* of the Inverted Earth sky looks to me just the same way as the *colour* of the Inverted Earth sky looked to me a long time ago.

b) The *blueness* of the Inverted Earth sky looks just the same then, as it does now.

Assuming that it is the same sky that it is being looked at, is there any grounds for doubting claim A? There might be good grounds for doubting claim B, but not because the way the colour of the sky looks to the Earthling has changed, but *only* because the meaning of 'blue' has changed. It is just not clear that the falsity of A follows from the falsity of B. It is, however, A that is maintained to be true by the proponent of the Inverted Earth argument, and it is not clear that the representationist has given any reason for doubting why A might be true. Now there might be good grounds for offering a general scepticism about the human ability to remember the colours of things, (this happens to be notoriously bad), but it is not clear that poor colour memories are going to account for a phenomenal re-inversion in an individual without that individual noticing.

Furthermore, just as the phenomenist need not appeal to transitivity of close shades to establish that there is no phenomenal re-inversion, it is not clear that the phenomenist needs to appeal to the memories of the individual to justify the unlikelihood of a phenomenal re-inversion. The phenomenist has independent grounds for arguing that there is no unnoticed phenomenal

---

[29] Block (1996) does indeed modify his scenario by making the Earthling aware of his lens insertion, but Block's motivation is to make it easier for the Earthling to become embedded in his new linguistic.

re-inversion. An appeal to empirical experiments do not support a phenomenal re-inversion because there are not any empirically discovered *relevant* changes in actual experiments that could be inferred to occur when the Earthling arrives on Inverted Earth and after he stays on Inverted Earth that could possibly warrantthe claim that he somehow compensates for his inverting lenses and starts to see the actual colours of objects on Inverted Earth. But there are other independent grounds for doubting slow unnoticed phenomenal re-inversion. The representationist who appeals to unnoticed slow phenomenal re-inversion needs to account for the fact that the phenomenal re-inversion has to follow an extremely specific path to remain unnoticed because of considerations concerning coextensive discriminations.

2.3.3.7 Coextensive discriminability again

There are arguably extremely good grounds for thinking that the Earthling will notice a phenomenal re-inversion because of a failure of coextensive discriminability as the slow inversion takes place. The representationist assumes that the slow phenomenal re-inversion will occur in a manner that maintains coextensive discriminability with all colours as the slow inversion occurs. But what possible grounds are there for *this* assumption?

I will illustrate with an example. Imagine that the Earthling is looking at a flag with a purple patch on top and a yellow patch on the bottom. The Earthling puts lenses on so that the top patch looks yellow to the Earthling and the bottom patch looks purple to the Earthling. Let us also assume that as the Earthling compensates for the lenses and undergoes a phenomenal re-inversion, the re-inversion is a slow creeping through the spectrum that follows a certain path. The purple slowly creeps through an indeterminate gray/brown and then to yellow. The yellow slowly creeps through an indeterminate gray/brown and then to purple. The problem should now be apparent. Halfway through the process the top of the flag will look the same as the bottom of the flag. Surely the perceiver will notice this. At some point violets will look to the Earthling, the same colour as bananas. There need not be a complete identity between the two colours for the change to be noticed, merely a converging of the two colours; the top of the

flag starts to look more like the bottom of the flag and the perceiver notices *this*.

The representationist might reply to the above point by claiming that the re-inversion of purple to yellow follows a path distinct from that of yellow to purple; thus coextensive discriminations between the two colours remains the same; i.e., the top of the flag looks just as different from the bottom of the flag throughout the whole process of re-inversion. But can the representationist really plausibly maintain that *all* the colours, as they pursue paths to their inverted opposites, follow paths that never intersect? The only possible colour model that the representationist requires is that of a colour circle. Furthermore, during the re-inversion each colour follows the other, like obedient ducks, either always clockwise or always counter-clockwise, around the circumferenceof the circle; for example, the bottom of the flag follows purple to blue to green to yellow; the top, contemporaneously follows yellow to orange to red to purple. Does the representationist have any plausible grounds for justifying this?

If purple goes to yellow, clockwise, and yellow goes to purple, counter-clockwise, then the representationist meets with a difficulty; the colours will meet at blue/green. If the re-inversion path goes through the center of the circle, across the indeterminate mixed shades of gray/brown, the representationist is also in trouble; the colours will meet at gray/brown. Furthermore, if the model we adopt is a colour *sphere* or a *solid*, then it is very unclear how it is supposed to be possible that *every* colour slowly changes, during phenomenal re-inversion, in a manner that allows each colour to end up on the location that corresponds to its inverted opposite on the colour solid without some of the re-inversion paths intersecting; or at the very least either converging or diverging. Phenomenal re-inversion in such a manner might be possible, but I would argue that such a re-inversion is extremely unlikely.

The representationist requires a phenomenal re-inversion without really offering any good grounds for why there should be a phenomenal re-inversion. One putative reason might be substantial self-motivated behavioural re-adjustment, but this does not apply to Inverted Earth. The

representationist also requires an *unnoticed* phenomenal re-inversion that maintains the coextensive discriminations of the Earthling as all the colours slowly phenomenally re-invert past one another. A total spontaneous re-inversion would arguably resolve problems of co-extensive discriminations, but a total and spontaneous re-inversion is hardly going to remain unnoticed by the perceiver. If the representationist has a reply to Inverted Earth, I suggest that they reject both slow creeping phenomenal re-inversions and total spontaneous phenomenal re-inversions.

2.4.1 Narrow contents

Earlier I suggested that there were two general strategies for the representationist to reply to Inverted Earth; the second is to argue that the Earthling actually compensates for his lenses and undergoes a phenomenal re-inversion and this corresponds with a change in intentional content. This reply is argued above to be substantially implausible, although it still currently has its adherents (Lycan, 1996, Tye, 1998a, 1998b, Lalor, 1999). The first strategy is to argue that the intentional content of the Earthling's mental states does in fact stay the same after the arrival of the Earthling on Inverted Earth. One way of maintaining intentional content constancy is to appeal to the narrow intentional contents of the Earthling's mental states. The narrow contents stay the same; the wide contents change (Rey, 1998, pp. 439-40).[30] Sydney Shoemaker (1998) presents the narrow content view as a possible reply to Inverted Earth, thus:

> Another way of dealing with the possibility of inversion . . . is to hold onto the claim that phenomenal character is representational content, but to abandon the view that content is externalist (p.674).

There is some confusion as to what constitutes a narrow content. It is easy to *over*emphasise the importance of physiological identity and local supervenience relations between mental states and their underlying physiology. Tye (1994) arguably fell to this temptation when he presented his defense of a narrow content account of phenomenal content or character. I will adopt a Fodorian type of conception of narrow content. I will assume

---

[30]Rey (1998) explicitly uses Inverted Earth as a successful refutation of wide representationism to justify narrow-content representationism.

that this is the best account of narrow content, i.e. if this account fails, no narrow content account of phenomenal character is adequate. It will eventually be argued that this account of narrow content does in fact fail when applied to phenomenal content or character. However, before the eventual demise of phenomenal narrow contents is demonstrated, the narrow content account will be developed independently of Inverted Earth and then it will be applied to Inverted Earth, thus explicating Shoemaker's quote. Finally, the flaws with a narrow content account when applied to phenomenal content will be demonstrated.

But what does it mean to claim that narrow contents are functions that map contexts onto contents? The context, within which the mental state's narrow content functions, determines the mental state's wide intentional content. The content of the mental state that is shared across distinct contexts is the narrow content. This conception of narrow content is best illustrated in terms of Fodor's response to Twin Earth.

2.4.2 Twin Earth, narrow contents and Inverted Earth

Twin Earth is a place very much like Earth except on Twin Earth water is not composed of H2O, but XYZ, henceforth to be referred to as twater. We are also asked to imagine that we are at a time where no one knows the chemical composition of water. On Twin Earth there is a community of molecularly identical humans, henceforth to be described as Twinlings. Twinlings also describe the stuff that flows in rivers and falls from the sky as 'water'. How do we evaluate the differences between the Earthling and Twinlings' beliefs about the substance that flows in the rivers and falls from the sky? At first blush, one might be inclined to argue that because the Earthlings and Twinlings are physically identical, they must have the *same* beliefs about the stuff that falls from the skies and flows in the rivers. On second blush, we might be troubled by the fact that the Earthlings' beliefs are about H2O while the Twinlings' beliefs are about XYZ. If beliefs are individuated by what they are about, then it appears that we have to infer that the Earthlings have *different* beliefs from Twinlings, concerning the substance that falls from the sky and flows in the rivers. How are these two claims to be resolved?

Fodor (1990) proposes an appeal to narrow contents as functions that map contexts onto truth conditions to resolve these two apparently incompatible claims. Imagine a scenario where the Earthlings and the Twinlings share a context, transport the Earthling to Twin Earth. If an Earthling and a Twinling share a context, for example Twin Earth, then the two individuals share a function which maps the same substance, in this case twater, onto the individuals' mental states. However, if the two individuals do not share a context, then their mental states will be mapped onto by different objects; and thus, the contents of their beliefs will have different truth conditions. They will have distinct wide contents, but shared narrow contents. Therefore, Fodor (1990) proposes narrow contents as functions that map contexts onto truth conditions:

> Well there is presumably something about the relation between Twin-Earth and Twin-Me in virtue of which his 'water' thoughts are about XYZ even though mine are not. Call this condition that is satisfied by me {Twin-Me, Twin Earth} condition C (because it determines a context). Similarly, there must be something about the relation between me and Earth in virtue of which my water thoughts are about H2O even though my Twin's 'water' thoughts aren't. Call this condition that is satisfied by {me, Earth} condition C1 . . . [I]t must be possible to satisfy C without satisfying C1 and vice versa . . . because . . . if an organism shares the neuro-physical constitution of my twin *and satisfies* C it follows that its thoughts and my thoughts share their truth conditions . . . in a world where I am in my Twin's context, given the neuro-physical identity between us, my 'water' thoughts are about XYZ if his are. . . . [T]wo intentions are identical only if they effect the same mapping of contexts onto truth conditions . . . [M]y Twin's 'water' thoughts are intentionally identical to my 'water' thoughts; they have the same intentional content even though their contexts are *de facto* different, they differ in truth conditions. In effect what we have here is an extensional criterion for what is sometimes called "narrow" content (pp.424-5).

The narrow content representationist, by applying this notion of narrow content to phenomenal character, offers a reply to Inverted Earth. They argue that the *narrow* intentional content of the Earthling stays constant after the Earthling arrives on Inverted Earth, has lenses inserted, and looks at the sky. Consider Tye (1994) once again:

> If perceptual experiences have narrow intentional contents . . . then there will be intentional aspects of your experiences which do *not* change through time, just as there are unchanging phenomenal aspects. The sky will continue to look blue to you, even if through time it comes to look to you the way yellow things normally look and you come to think of it as being yellow (p.172).[31]

When Tye here claims that through time the sky will look to you the way yellow things normally look, he is *not* claiming the sky will start looking yellow to the Earthling. He is claiming that the wide content of your mental states will change as your mental states start to co-vary with the environment of Inverted Earth. However, because, according to the narrow representationist account, your experiences have narrow intentional contents, you can get around Inverted Earth, by claiming that these narrow intentional contents do not change in the Earthling after the move to Inverted Earth and however long the Earthling stays on Inverted Earth. The Earthling's mental states are the *same* mental states, only in a different context. The different context maps onto the same mental state, thus providing a distinct wide content, but it is the same mental state with the same narrow intentional content.

We might look at the specific neurophysical constitution of the brain of the Earthling before and after the move to Inverted Earth. The brain goes into the same states when presented with the same types of objects on Inverted Earth. This is equivalent to the 'water' beliefs of an individual who travels to Twin Earth. However, the wide contents of the Earthling's mental states change when he goes to Inverted Earth, just as the individual's 'water' belief is about XYZ on Twin Earth.

This representationist reply takes on board the intuition that the phenomenal content or character of the Earthling who travels to Inverted Earth stays the same. This constancy of phenomenal content or character is explained by *identifying narrow contents with phenomenal content or character* and maintaining that the narrow contents of the Earthling stay the same after the move to Inverted Earth. This representationist reply also takes on board the intuition that there is a change in the intentional content of the Earthling's

---

[31]Tye (1995b) rejects narrow contents.

mental states. There is a change in the *wide* intentional content of the Earthling's mental states. Therefore, according to this reply, the Inverted Earth argument fails to establish that phenomenal content or characteris not entirely constituted by intentional content. However, this reply does fail to establish that phenomenal character is entirely constituted by wide intentional content.

There are, however, independent grounds for rejecting a narrow content account of phenomenal character.

2.4.3 Undermining the narrow content account of phenomenal content or character

The argument that undermines the narrow content account of phenomenal characterfollows a similar strategy as the Inverted Earth argument. I will attempt to demonstrate the possibility of two mental states having distinct narrow intentional content whilst they have the same phenomenal contents or character. If this is possible, we have to conclude that narrow intentional content does not entirely constitute phenomenal content or character. Narrow contents are not phenomenal contents.

When considering close shades of colour it is possible that narrow contents might be considered to be the same in certain contexts due to phenomenal indistinguishability, but the phenomenal content of one of these shades might also vary depending on the context under which this shade is viewed thus suggesting that narrow contents should not be identified with phenomenal content.

Consider four very close, but distinct, shades of blue: b1, b2, b3 and b4. An individual is capable of discriminating between shades that are more than one shade apart. This individual can distinguish between b1 and b3, and between b2 and b4. However, this individual is not capable of distinguishing between b1 and b2, b2 and b3 and b3 and b4. Given the fact that this individual is not capable of distinguishing between b2 and b3, it will be assumed that when the individual is looking at a patch of paint constituted by b2 and a patch constituted by b3, the phenomenal content or

characterof his mental states is shared. The two patches 'look the same' to the perceiver. The narrow content theorist is committed to explaining this phenomenal similarity by an identity between the mental state's narrow content 'b2' and the mental state's *same* narrow content 'b3'.

At first, this argument appears to *support* a narrow content account of phenomenal content or character. Consider the *wide* contents of this individual's mental states when he is confronted with first a patch of b2, then a patch of b3. Remember narrow contents are functions that map contexts onto contents. When this individual is confronted with a patch of b2, this is a distinct context, from when he is presented with a patch of b3. His mental state clearly has the *wide* content b2 and not b3. When this individual is presented with a patch of b3, this is a distinct context from when he is presented with a patch of b2. His mental state clearly has the *wide* content b3 and not b2. But, *are* the narrow contents 'b2' and 'b3' shared? Crucially, are the narrow contents 'b2' and 'b3' shared across other distinct contexts? The answer appears to be, no.

Now consider when the individual is presented with a patch of b2 and a patch of b4. The narrow content of his mental state 'b2' is a function that maps a context that enables the individual to differentiate the mental state with the narrow content 'b2' from 'b4'. Now consider when the individual is presented with a different context, a patch of b3 and b4. The narrow content 'b3' is a function that fails to differentiate the narrow content state 'b3' from 'b4'. This suggests that the narrow content state 'b2' that maps b2 is *distinct* from the narrow content state 'b3' that maps b3. The function is not shared because the narrow contents function differently, across the two contexts when the two distinct narrow contents are placed in these distinct contexts. Therefore, the narrow contents 'b2' and 'b3' must be distinct.

Consider when this individual is presented with a patch of b1 and b3. Likewise, this individual has a mental state that maps a context whereby narrow content 'b3' can be differentiated from 'b1'. Now consider when the individual is presented with a distinct context: a patch of b1 and b2. The narrow content of the mental state 'b2' that maps b2 cannot differentiate

between 'b1' and 'b2'. This also suggests that the narrow content 'b2' that maps b2 is distinct from the narrow content 'b3' that is maps b3. Not only does the narrow content of the mental states that are map b2 and b3 differ in *wide* contents, there is *nothing shared* by these two narrow contents when they are placed in distinct contexts.

Narrow content 'b2' is *distinct* from narrow content 'b3,' when they are placed in different contexts, *yet* the phenomenal characterof the mental state of an individual when presented with b2 is identical to the phenomenal character of a mental state when presented with b3. Therefore, narrow content does not entirely constitute phenomenal content or character.

2.4.4 What are the possible narrow content replies to this argument?

The narrow content representationist might reply to the above argument by claiming that the phenomenal content or character also changes across different contexts. But this does little to explain why the phenomenal content or characterof mental states when exposed to patch b2 and b3 is the same when presented together. It is not because the narrow content 'b2' and 'b3' is shared, as the narrow content theorist would like to suggest.

The narrow content representationist might try to *deny* that the phenomenal content or characterof the mental states when presented with patch b2 and b3 is shared. The phenomenal content or character of the mental states when presented with patch b2 and b3 are distinct; and hence, the narrow contents are distinct. This entails that the person does not have introspective access to the differencebetween these distinct narrow contents. The narrow content theorist is required to deny the introspective accessibility of distinct narrow contents, and hence argue that the phenomenal content or characters of the individual's mental states are distinct. The person does undergo distinct phenomenal content or characters when presented with b2 and b3, the person just does not realize it. However, this would be a most curious, arguably inconsistent, line of defense for a *narrow* content theorist to take.

If the *narrow* content theorist argues for distinct phenomenal content or character, it seems that the narrow content theorist is forced into denying comparative self-knowledge of content. But this appears to be one of the major motivations for narrow contents over wide contents in the first place. If this is denied, then there is nothing to stop the *wide* content theorist from denying the plausibility of Inverted Earth in the first place on the grounds that the Earthling does undergo distinct phenomenal content or characters on Inverted Earth, he just does not realize it. It is precisely comparative self-knowledge of content that motivates the claim that the Earthling's phenomenal content or character of his mental states stays constant after the move to inverted earth, which in turn motivates the narrow content reply to Inverted Earth. If the narrow content theorist rejects comparative self-knowledge, the main motivation for positing narrow content in the first place is undermined.

The phenomenal content or character of an individual's mental states when presented with two close shades of colour is the same. If the narrow content theorist attempts to deny this, he essentially underminines Inverted Earth, the motivation for narrow contents in the first place. However, the narrow contents of the mental states that correspond to 'b2' and 'b3' are distinct because these contents are not *shared* functions that map distinct contexts onto the contents. When placed in distinct contexts these narrow contents behave in very different ways. *Therefore, narrow content does not entirely constitute phenomenal character.*

Notice that this argument does not rule out wide content representationism, it merely serves to remove one representationist contender from the field. It is perfectly possible for the wide representationist to deny self-knowledge of mental states, indeed it might even be required, but this is not an option for the narrow representationist. Ruling out narrow representationism becomes increasingly important when considerations of causation are developed in part 2 because it is arguably the narrow content view that attempts to reconcile problems of causation with intentional content. Without an appeal to narrow content, considerations of causation and explanation become especially sharp against the representationist.

2.5 Are the fine-grained functional differences relevant?

It might be argued that the Earthlings are never completely functionally isomorphic with the Invertlings because of fine-grained functional differences between them.

> A parallel problem arises in connection with the earlier example of blackish blue versus blackish yellow. Clearly some modifications are needed in the Inverted Earth hypothesis for it not to encounter immediate difficulties (Tye, 1995b, p.206).

Arguably the failure of coextensive discriminability is perhaps the more serious of the fine-grained functional differences; the problem of blackish-yellow is a variant on this same point, i.e. black is closer to blue than to yellow. Indeed these fine-grained functional differences are relevant. I have employed them to cast doubt on the possibility of unnoticed slow phenomenal re-inversions. But the question remains as to whether they scupper the Inverted Earth argument. As pointed out earlier, it is not clear as to how fine-grained functional differences are relevant to a 'semantic' style of representationism. However, it is arguably the 'semantic' representationist who has the most to gain by pointing out functional differences between the Earthling and the Invertlings, but only if this difference has an impact on the meaning of their colour terms. It is simply not clear how these fine-grained differences are relevant to the meanings of colour terms.[32] Are the fine-grained functional differences relevant to the current long-arm functional representationist? Arguably not.

The reason that the fine-grained functional differences appear to be of no relevance to the current long-arm functional representationist is because the proponent of the argument does not require the presence of Invertlings to establish a long-arm functional difference between the Earthling's mental states before he arrives on Inverted Earth and after he arrives on Inverted Earth. The Invertlings might be useful to illustrate heuristically the change in long-arm functional roles of the Earthling's mental states, and hence the

---

[32] See Block (1999) for elaboration of how fine-grained functional differences are not relevant to this semantic style of representationism.

intentional content of the Earthling's mental states, but the Invertlings are not required.

It is precisely the long-arm functional differences between the Earthling before he arrives on Inverted Earth and after he arrives on inverted Earth that need to be exaggeratedand not diminished by the proponent of the Inverted Earth argument. With these functional differences comes a difference in intentional content. But as long as the phenomenal content or characterof the Earthling's mental states stay the same, then this type of Rerpesentationism apparently faces a problem. An appeal to fine grained functional differencesis not going to help this type of representationism.

2.6 Summing up Inverted Earth *thus far*

A semantic style of representationism that emerges out of 'anti-private language' considerations against inverted spectrums is one motivation for Inverted Earth. This style of representationism is rejected because it is not clear as to how the semantics of colour terms has anything to do with the phenomenal character of mental states. Phenomenal re-inversion, as one attempt to save both the semantic representationism and the current long-arm functional representationism is rejected for the following reasons: first, because one motivation for arguing for phenomenal re-inversion is inapplicable to Inverted Earth; and second, because of the presence of insurmountable problems concerning *unnoticed* phenomenal re-inversions. An appeal to narrow intentional contents is a good attempt to maintain intuitions behind the plausibility of both premises of the Inverted Earth scenario and that phenomenal content or characteris still entirely constituted by intentional content, but phenomenal narrow contents ultimately fail because of the 'four shades of blue' objection. Fine-grained functional differencesare simply not relevant in a manner that the representationist can use to justify a response to Inverted Earth. However, the representationist still has replies to Inverted Earth available, one more plausible than the other.

The next chapter considers the plausibility of a representationism that appeals to the historical wide functional role of mental states, sometimes

described as teleological accounts of intentional content. These will eventually also be deemed problematic as a plausible account of phenomenal content or character. The teleological representationist response to inverted Earth is rejected partly because of the availability of a far simpler but overlooked reply to Inverted Earth on behalf of the wide representationist. Finally a wide representationist reply to Inverted Earth will be considered that is superior to all the other replies; thus Inverted Earth will also be committed to the dust heap of mere metaphysical speculation, despite being a fascinating exercise in philosophical imagination.

Chapter 3: Inverted Earth (II)

3.1 Introduction

According to the proponent of the Inverted Earth argument, Inverted Earth demonstrates that the intentional content of an Earthling's mental states changes when the Earthling travels to Inverted Earth, yet the phenomenal content or character of the Earthling's mental states does not change; therefore, representationism is false.

An appeal to a slow or immediate phenomenal re-inversion in the Earthling's mental states after he arrives on Inverted Earth is an attempt to show that the phenomenal content or character of the Earthling's mental states do not stay the same; therefore, the above argument is unsound. However, in chapter two, it was demonstrated that an appeal to a phenomenal re-inversion of the Earthlings mental states is not a good strategy for the current long-arm functional representationist to adopt as a response to Inverted Earth; however, no independent grounds were offered for rejecting a current long-arm representationist account of phenomenal character. If, as will be demonstrated in this chapter, Inverted Earth fails on distinct grounds, then the phenomenist fails to demonstrate the falsity of this variant of wide representationism. Although narrow content representationism offers a creative solution to Inverted Earth, independent grounds were offered in the last chapter to demonstrate that narrow contents as functions that map contexts onto contents do not entirely constitute phenomenal content or character.

In this chapter, two more representationist replies to Inverted Earth are considered, the second more successful than the first. The first representationist reply applies to mental states a historical long-arm functional role account of intentional content. This representationist account appeals to the histories of mental states to determine the intentional content of those mental states. The current long-arm functional account of intentional content appeals to the current inputs and outputs of mental states and the way they function in a current environment to determine the

intentional content of the mental states. The historical account looks to the way the mental states functioned in the past to determine the intentional content of those mental states. The appeal to historical functional roles sub-divides into two further types of representationism as follows: phylogenetic representationism and ontogenetic representationism. But although historical accounts of intentional content are plausibly applicable to non-phenomenal mental states, problems concerning their applicability to phenomenal states will be discussed.

The second representationist reply to Inverted earth that is considered in this chapter makes no appeal to historical long-arm functional roles, nor does it appeal to narrow contents. This reply makes no appeal to a semantic style of representationism motivated out of anti-private language considerations and it certainly does not appeal to phenomenal re-inversion in the mental states of the Earthling after he arrives on Inverted Earth. The successful reply, however, brings out some essential ambiguities in the presentation of the Inverted Earth argument that can be successfully exploited by the representationist.

3.2 History, teleology and content

The histories of mental states of the Earthling before he arrives on Inverted Earth and even before he was born are appealed to in an attempt to argue that there is something about this history of the mental states of the Earthling that *fixes* the intentional content of the Earthling's mental states. If the intentional contents of the Earthling's mental states are *fixed* in the past of the Earthling, then there is no reason to maintain that the intentional content of the Earthling's mental states changes, after he arrives on Inverted Earth; therefore, the Inverted Earth argument is unsound. If one considers the Invertlings, there might be something in the history of the Invertlings that fixes the intentional content of their mental states and if their history is significantly different from the history of the Earthlings, then according to a historical account of intentional content, there is no reason to maintain that the intentional content of the Earthling's mental states ever comes to resemble the intentional content of the Invertling's mental states; this

undermines any purported support for a change in the Earthling's intentional argument; therefore, the Inverted Earth argument is unsound.

The relevance of the histories of our mental states to intentional content is arguably motivated by certain intuitions concerning how representational states acquired their contents *in their past*. For example, photographs are about what they are about because of events that occurred in the history of the photograph. Arguably what determines the content of the photograph is the process that occurred in the past whereby light bounced of an object, travelled through a lens and exposed a negative. It is this event that occurred in the past that determines the intentional content of a photograph and nothing to do with the 'current' functioning of that photograph. Although the photograph may well be used to identify the object originally photographed, this is not relevant to the intentional content of the photograph. According to a historical account of content, the intentional content of a mental state is determined by events that occurred in the history of that mental state.

However, according to teleological accounts of intentional content it is not *only* the history of the mental state that matters, but what the mental state was 'designed' to represent that determines the intentional content of the mental state; hence these accounts are often described as 'teleological'. Stephen Stich (1990) offers a good summary of Fred Dretske's use of the notion of a 'designed' state:

> A simple example is the bi-metallic strip in a thermostat, whose function is to indicate the temperature. When embedded in the thermostat, the bi-metallic strip not only indicates the temperature, it is also a *representation* of the temperature, since, for Dretske, "a representational system" is "any system whose function it is to indicate how things stand with respect to some other object or magnitude." "What a system *represents* . . . is what these elements have the *function* of indicating . . . In the case of the thermostat, it is the person, the designer of the thermostat, who determines what the function of the indicator will be. But Dretske maintains there are also natural biological systems in which internal indicators have acquired the function of indicating something, typically something about the organism's external environment (p.802).

Two distinct theories of teleological content pull apart depending on which causal histories are deemed to be relevant for attributing the 'design' of the mental state. Dretske (1981) maintains an ontogenetic teleological account whereby learning and conditioning in the specific history of *that* mental state determines its function and hence its intentional content. phylogenetic teleological accounts maintain that the evolutionary history of the individual and the environmental pressures that selected for the mental state determines the design and historical function of that mental state; and hence, its intentional content. This distinction is important because differing variants on the teleological accounts are sufficient for differing responses to Inverted Earth.

According to ontogenetic accounts of intentional content, the design history of mental states that is relevant to the function of those mental states is the development of individual mental states as a result of conditioning and learning (Dretske, 1981). Stich (1990), although ultimately disagreeing with Dretske, offers an excellent summary of Dretske's version of an ontogentic teleological account of content with the example of conditioning in a rat:

> In conditioning, according to Dretske, an internal indicator, C, comes to be causally connected to another internal state which, in turn, causes the organism to produce a certain sort of behavioural output, M. Moreover, C gets hooked up to M - it gets to be a cause of this sort of behavioural output - *because* M is an appropriate way to behave when the state of affairs that C indicates obtains. Thus when C gets linked to M, it comes to be C's function to indicate that state of affairs. An example will make this a bit clearer. Consider the rat in the Skinner box. When the light is on, pressing the bar will produce a bit of food. Let F be the state of affairs in which the light is on. Presumably there is some internal perceptual state of the rat that indicates when the light is on. Let this state be C. Now prior to being placed in the Skinner box, C was not causally linked up to the bar pressing behaviour (=M). But as a result of conditioning, a causal link between the two was forged. C comes to cause M because it indicates F. It is now C's function to indicate F, and as we've seen, for Dretske an indicator of F whose function is to indicate F is a *representation* of F (p.802).

So for Dretske (1981), a natural system's way of utilising indicators in order to acquire the function of indicating something is through learning

and conditioning. For a phylogenetic representationist, the natural selection of individuals within a species determines the biological function of the mental states.

According to phylogenetic accounts of intentional content, the design history of mental states is also relevant, but not a history that involves the learning processes as an individual mental state develops in an individual as a result of conditioning and learning, but a history that involves the selection pressures on the mental states of an individual as its ancestor's evolved. The phylogenetic account of the intentional relation between our mental states and objective properties in our environment emphasises the selection pressures on the survival of the ancestors of the owner of those mental states; the selection pressures determine the biological functions of those mental states. According to the phylogenetic representationist, the history of the ancestors of the owner of the mental states, more specifically, the selectional history of the individual's mental states determines the historical function of the mental state; and hence, the intentional content of mental states. Therefore, the evolutionary history of the species of the organism that possesses the mental states is salient to determining the intentional content of those mental states. Papineau (1998) explains the teleological theory thus:

> The central thesis of the teleological theory . . . is that the representationalcontents of mental states can be explained in terms of the biological functions of those mental states. More specifically, this theory equates the contents of informational states . . . with the conditions in which those states are biologically supposed to be present. . . . Defenders of the teleological theory standardly cash out their references to biological purposes *etiologically*, in terms of processes of natural selection. According to the etiological account of biological function, as item X has a biological purpose Y if and only if X is now present because previous versions of X were selected in virtue of doing Y. The paradigm case of such selection is inter-generationalDarwin selection of genetic traits (p.1).

This account is best illustrated with examples. Most humans have mental states that enable them to discriminate between different objects by virtue of their colours, for example, ripe from raw tomatoes. This might be explained in terms of the survival advantage conferred on a certain proportion of a previous population who are able to discriminate between

objects on the basis of differences in colour. The ability to discriminate between objects of different colours may have come about by genetic mutation or a recombination of genes, genes that are responsible for the development of cones in our eyes. Natural selection will then select for the proportion of organisms that inherit genes responsible for discriminable capacities in organisms based on colour. Therefore, natural selection explains why we have certain mental states that enable us to discriminate between objects on the basis of colour. The phylogenetic representationist then applies this account of natural selection to explain the biological function of colour detecting mental states.

According to the phylogenetic representationist, selection pressures in the past of our mental states determine the function of our colour indicating mental states. Evolutionary processes design our mental states and lock our mental states onto the features of the natural environment that we once occupied in our past. Certain mental states have colour as their intentional content because colour detecting is the biological purpose of those mental states. Mental states that detect colour are only present in us now because previous versions of colour detecting states were selected in the past in order to detect colour. [33]

Another application of the account is the ability of frogs to detect flies. Arguably selection pressures favour frogs that can catch flies; therefore, the frogs that possess mental states that recognise flies will be more likely to pass on the genetic trait that enables the frog to catch flies. Crucially, however, what is important to the phylogenetic representationist is not how frogs come to be able to catch flies, but why the mental states of the frogs represent flies. The frog's mental states *are about* flies because of the history of the fly detecting mental states of the frog; more specifically, evolutionary pressures favoured a fitness advantage for the frog's ancestors whom possessed the genetic trait responsible for these mental states; thus designing the frog's mental states to detect flies. Certain mental states in

---

[33] An immediate problem arises with this account because of an ambiguity between trait-types and trait-tokens. If natural selection only offers an explanation of trait-types or trait frequencies within a population, then it is not clear that a phylogenetic account will able to offer natural selection as part of a causal history as to the development of a token mental state in one individual (Walsh, 1998).

frogs have flies as their intentional content because fly detecting is the biological purpose of those mental states; fly detecting mental states are only present in frogs now because previous versions of fly detecting states were selected in the past in order to detect flies.

Having offered a brief summary of two teleological accounts of intentional content, it will be demonstrated as to how this is relevant to Inverted Earth and problems with these accounts will be demonstrated.

3.3.1 Teleology and Inverted Earth

When the Earthling arrives on Inverted Earth, his mental states act in a manner for which they were not originally designed. Mental states previously used to identify red tomatoes as red now identify green tomatoes as red. But according to the teleologist, it is not the current long-arm functional roles of these mental states that determines the intentional content of these mental states; it is what these mental states were originally designed to do that fixes the content of these mental states. Tye (1998a, pp.679-81) gives a parallel example of an axle speedometer being shifted to a planet where in one sense the speedometer 'works' but it fails to fulfil the function it was originally designed to fulfil.

Imagine on a new Twin Earth, an axle speedometer from a car on Earth is installed on a car on Twin Earth. On Twin Earth distances are measured differently and all the wheels of cars are systematically of a different size: these differences are such, however, that they cancel out. From a Twin Earthling's perspective the speedometer functions perfectly adequately. If the speedometer were to read '50' on Twin Earth, this would misrepresent the speed in terms of Earth miles per hour because the wheels are the wrong size; this malfunctioning is compensated by the fact that Twin Earth 'miles' are a different length, so the speedometer appears to function adequately to the Twin Earthling's on Twin Earth. The Dretskean and Tyean intuitions are that the speedometer *misrepresents* the speed of the car on Twin Earth because *it was designed to measure Earth velocities*, despite the fact that it functions adequately from a Twin Earthling's perspective. Notice that a molecularly identical Twin Earth axle speedometer, designed for cars on

Twin Earth, represents Twin Earth velocities accurately because that is what the Twin Earth axle speedometer was designed for. Applying the Teleological account of content, design determines content; therefore, the Earth speedometer misrepresents whilst the Twin Earth speedometer represents speeds accurately.

According to the teleologist, the axle speedometer is analogous to the colour detecting mental states of the Earthling. The mental states of the Earthling were designed to detect colour on Earth. But by a quirk of 'fate', strange colours and inverting lenses, the mental states of the Earthling *appear* to function adequately on Inverted Earth. When compared to the mental states of the Invertling, the mental states of the Earthling appear to both the Invertling and the Earthling to function adequately just as the Earth speedometer appears to function adequately on Twin Earth. However, just as speedometers are not designed for use on Twin Earth, the mental states of the Earthling are not 'designed' for use on Inverted Earth; therefore, the mental states of the Earthling *malfunction* on Inverted Earth. The intentional content of the Earthlings mental states cannot be compared to the mental states of the Invertlings because their mental states have different design functions. According to teleological accounts of content, design function determines intentional content. Furthermore, after the Earthling travels to Inverted Earth, his mental states fulfil the same design function that they were designed to fulfil, albeit malfunctioning on Inverted Earth; therefore, the Earthling's history permanently fixes the intentional content of the Earthling's mental states.

The Earthling's mental states were designed to identify, for example, red objects as red. When he arrives on Inverted Earth, this is not his natural habitat. His mental states are performing an artificial function that only appears to work from both the Earthling and the Invertlings' perspective, just as the axle speedometer appeared to work when it was installed on a Twin Earth car. However, because the mental states of the Earthling are fulfilling a role for which they were not *designed*, there is no reason to suppose that the intentional content of the Earthling's mental states are either identical to the Invertlings mental states or that they change.

> If what an experience normally tracks is what nature designed it to track, what it has as its biological purpose to track, then shifting environments from Earth to Inverted Earth will make no difference to normal tracking and hence no difference to the representational contents of your experiences (Tye, 1998b, p.461).

According to the teleological accounts of content, the history of the Earthling's mental states fixes the intentional content of the Earthling's mental states; thus no new environment changes the relevance of the design function of the Earthling's mental states. The intentional contents of the Earthling's mental states do not change; therefore, the Inverted Earth argument is unsound.

Thus far, the representationist reply only appeals to design functions, but remains ambiguous as to whether a phylogenetic accounts or an ontogenetic account establishes the relevant design history that determines the intentional content of mental states; and hence, the phenomenal content or characterof the Earthling's mental states.

3.3.2 Phylogenetic teleology applied to Inverted Earth

According to the phylogenetic representationist, the Earthlings have an evolutionary history that determines the intentional content of their mental states. This locks the intentional content of the Earthling's mental states onto the properties of the Earth environment in the past of the Earthling. The Earthling can never overcome his evolutionary history; therefore, his intentional contents will always represent what they were designed to represent. Moving the Earthling to Inverted Earth is like moving a fish out of water except the Earthling can still breath and function 'normally'. His mental states will never come to represent the colour of objects on Inverted Earth because this scenario was not a scenario for which his mental states were originally, in an evolutionary sense, designed.

> I am a living creature with an evolutionary history; the environment in which I find myself on Inverted Earth is not my natural habitat and I am wearing inverting lenses. So, on Inverted Earth, optimal conditions do not obtain. The brain state that tracks blueness in optimal conditions (and thereby represents blueness) now tracks yellowness. But it does not now *represent* yellowness (Tye, 1995b, p.205).

> The sensory state that nature designed in your species to track blue in the setting in which your species evolved will continue to do just that even if through time, on Inverted Earth, in that alien environment, it is usually caused in you by looking at yellow things (Tye, 1998a, p.461).
>
> I understand teleology in terms of selection history . . . thus for me, the intentional states of the victim transported to Twin Earth would *not* change their intentional contents even if that person stayed on Inverted Earth for a very long time and the states ceased having their evolutionary normal causes altogether (Lycan, 1996, p.114).

According to the phylogenetic teleologist, the intentional content of the Earthling's mental states never changes, so the phenomenal content or characterof the Earthlings mental states never changes. The Inverted Earth argument is thus no threat to the historical long-arm functional representationistwho appeals to the phylogenetic history of mental states.

3.3.3 Applying ontogenetic teleological representationism to Inverted Earth

Ontogenetic teleological representationism might also be appealed to, in order to respond to Inverted Earth on behalf of the representationist. It is, however, worth distinguishing another sub-division, between two types of ontogenetic teleological representationism as follows: The first, assumes that all developmental learning and conditioning ends at the end of child development; thus all intentional content of mental states is fixed at the end of child development. This position might be described somewhat clumsily as developmentally fixed ontogenetic teleological representationism. If this is so, then all the intentional contents of the Earthling's mental states are fixed by the time he has reaches a certain age. These intentional contents are fixed by the Earthling's Earth environment. When the Earthling goes to Inverted Earth, the intentional content of all his mental states have been 'designed' by ontogenetic development -- conditioning and learning throughout the Earthling's youthful, formative years *on Earth*. The Earthling's mental states permanently stay the same after he arrives on Inverted Earth; therefore, the Inverted Earth argument

presents no challenge to the developmentally fixed ontogenetic teleological representationist.[34]

The second type of ontogenetic representationism allows for continuous development, learning and conditioning throughout an individual's life. This position might be described as developmentally *continuous* ontogenetic teleological representationism. However, on this account, the above teleological strategy of maintaining that the intentional contents of the Earthling's mental states remain constant after the Earthling arrives on Inverted Earth cannot be employed. This position arguably most closely resembles the current 'long-arm' functional account of mental states. The short-term history of learning and conditioning fixes the intentional contents of these mental states, these may take some time to establish whilst a creature adapts to its new environment. Given continuous development and adaptation to a new environment between the mental states and the properties of the new environment, the intentional content of the Earthling's mental states changes on Inverted Earth. This therefore commits the developmentally continuous ontogenetic teleological representationist to the Earthling undergoing a phenomenal re-inversion. Phenomenal re-inversion is not, however, a plausible option for the representationist, given the Inverted Earth scenario (see last chapter); therefore, representationism of *this type* is not going to offer a coherent reply to Inverted Earth.

But, how about the phylogenetic and the developmentally fixed ontogenetic representationist replies to inverted Earth? Are they successful? These representationist theories, however, also have problems associated with them independent of Inverted Earth.

3.4.1 Problems with phylogenetic and developmentally fixed ontogenetic teleological accounts of intentional content as entirely constituting phenomenal content or character

---

[34]It is perhaps worth pointing out that Inverted Earth might present a challenge to this type of representationism if the Earthling is taken to Inverted Earth as a young child that is still undergoing the required development that fixes the child's intentional contents.

Problems with teleological accounts arise because they have problems accounting for the specific occurrent contents of our mental states. The specific occurrent contents of mental states are the specific contents of the mental states of a perceiving organism contemporaneous with the organism being presented with an object in its environment at a determinate point in time. According to the phylogenetic account, the explanation for the specific occurrent content of our mental states is not properties of our immediate environment, but a long history of selective pressures on our mental states in prior ancestors. But is this really plausible? Certainly it does seem plausible to claim that our perceptual systems have evolved in a certain manner; human eyes have evolved in very different manner from insect eyes. However, it is not clear that selection pressures should apply to the specific contents of those mental states.

One possible teleological claim that does perhaps have some plausibility is that selection pressures might select for the kinds of properties that enter into the content of our mental states. So, for example, properties of depth perception might be more important concerning selection pressures to certain creatures while detecting movement might be more important to other organisms. But the relevance of selection pressures on the kinds of properties detected by our visual systems does not entail that the selection history of our perceptual system determines the specific occurrent contents of our mental states. It is a particular problem to teleological accounts of colour that the survival functions for specific colour experiences come out the same providing discriminatory capabilities come out the same. Surely the most obvious explanation for specific phenomenal contents is the properties of the immediate environment of a perceiving organism. Two additional metaphysical speculations might be proposed to demonstrate that the current functions of mental states are better suited to determining the intentional content of the mental states that putatively entirely constitute phenomenal content or characteras follows: the first is swamp-creatures and the second is the silicon replacement of brain parts.

A further problem with historically based accounts of intentional content is that they rarely take into account the possibility that mental states have evolved to function in new environments. Arguably our perceptual systems

have not just evolved to function in one specific environment. They have evolved to function in any new environment that we happen to occupy. This is a somewhat useful feature of perceptual systems. We can enter environments that we have never occupied before and survive and thrive in these new environments. Consider an organism leaving its natural habitat for the first time. According to the evolutionary account given above, this creature will not be representing correctly because it is entering an environment for which the specific contents of its mental states have not been selected for, but surely this is not true. When we enter new environments, new properties enter into the contents of our perceptual systems. Our perceptual systems are designed so that occurrent environments can fix the phenomenal content or character of our mental states.

One further problem specific to the phylogenetic account is that it is not clear as to how belonging to a species is specified. Imagine if the Earthling marries an Invertling and has a child on Inverted Earth. Let us assume that it was a genetic alteration that inverted the colour experiences of the Earthling and not inverting lenses.[35] This genetic alteration is then passed on to the child and all the future offspring of that child. Arguably being born on a planet is not sufficient for belonging to a species. So the child is not yet an Invertling. Although having half the genes of an Invertling might suggest it is on the way to becoming an Invertling. In the future perhaps we will say that the descendents of the Earthling are Invertlings, but how will we then describe the contents of those descendents' mental states? It is certainly not obvious to me that the answer should be any different from the description of the intentional content of the mental states of the earthling with the genetic alteration when he first arrived on Inverted Earth.

Before a superior reply to Inverted Earth is offered, these objections to teleological accounts will be developed further in the next few sections, not as a knock down refutation of these theories, but as a serious questioning as to whether teleological accounts of intentional content are really applicable to phenomenal content or character. Inverted Earth is eventually going to be

---

[35]Perhaps the genetic alteration affects the cones of the Earthling in an identical manner to colour inverting lenses.

shown to fail as a refutation of representationism on grounds independent of teleological considerations. These independent grounds demonstrate that an appeal to a teleological account of phenomenal character is not required as a response to inverted Earth, and if Inverted Earth is an important motivation for adopting a teleological account of phenomenal content, then the existence of a better response to Inverted Earth is troubling to the teleologist. However, it is still worth developing some of these independent objections to a teleological account of phenomenal content or character in the next few sections.

3.4.2 Swamp creatures

Swamp creatures are just one example of a thought experiment that demonstrates that teleological accounts ignore the relevance of specific properties of an environment that immediately constitute the phenomenal content or character of the mental states of a perceiver. Swampman is a molecular duplicate of an individual that accidentally pops into existence. This qualifies as metaphysical speculation of the sort of which Gwendolen is so scornful. However, it appears that the existence of such creatures offer counter-intuitive implications if phylogenetic representationism or developmentally fixed ontogenetic teleological representationism is assumed

First, if an evolutionary history involving all the nitty-gritty of natural selection within a natural environment is a necessary condition for a mental state having *any* intentional content, then swamp creatures, despite having all the right 'bits', have mental states with *no* intentional content. The swamp creature has mental states that are of nothing and thus these mental states have no phenomenal content of character.[36]

---

[36] A teleological account might be modified thus: if there is no evolutionary 'design,' there might still be the possibility of the mental states having content, but a content of a different type. Tye (1998a,b) presents a dual account whereby swamp-people have phenomenal contents, but when a swampman has lenses inserted and is transported to Inverted Earth he phenomenally re-inverts; developmentally continuous ontogenetic teleological representationism applies to swampmen. For humans, design trumps current functioning, so the Earthling's do not phenomenally re-invert when they go to Inverted Earth; phylogenetic Rerpesentationism applies to humans. Arguably, however, Tye doubles his problems rather than solves them. Phenomenal re-inversion is problematic (see chapter two), even for swamp-people, and phylogenetic representationism faces problems independent of swamp-people (see rest of chapter).

> [Assuming phylogenetic representationism] accidental replicas of actual sentient creatures lack all experiences. . . On a cladistic conception of species, swampman is not human. Indeed lacking any evolutionary history, he belongs to no species at all. His inner states play no teleological role. Nature did no design any of them to do anything. So, if phenomenal character is a certain sort of teleo-representational content, then swampman has no experiences (Tye, 1998b, p.461).

Second, swamp creatures also present a problem to developmentally fixed ontogenetic Teleological representationism because a spontaneously created *adult* swampman does not have mental states that have gone through the appropriate development in order to create any intentional contents of these mental states. [37]

However, it is extremely counter-intuitive to suppose that such duplicates undergo no phenomenal experiences. For all anyone knows, anyone could turn out to be a swampman; yet arguably all of us undergo a rich and full experiential life. Surely kicking a swampman is not equivalent to kicking a football with regards to the pain felt. Therefore, according to the swampman objection, both phylogenetic representationism and developmentally fixed ontogenetic teleological representationism are false.[38]

The Teleological representationist might allow that the swampman undergoes some phenomenal character, but even then, the intentional contents of the swamp-man must be substantially *indeterminate.*

> A creature without any history or evolutionary ancestry would have *indeterminate* qualitative states, no matter how much it might physically resemble a normal person possessed of determinate such states . . . when swampNed looks at either a red tree or a green one - well, there is presumably no fact of the matter about what he is experiencing in either case (Rey, 1998, pp.443-4).

---

[37] Notice that the phylogenetic teleological representationist seemingly entails the possibility of non-feeling swamp babies growing into *non-feeling* swamp adults. Whilst the developmentally fixed ontogenetic teleological representationism can accept that the swamp-baby will develop mental states with intentional contents and phenomenal characters just like any other normal adult. These two teleological accounts are therefore inconsistent.

[38] A developmentally *continuous* ontogenetic teleological representationist has a response to swamp-people because as soon as they start to interact with their environments a new causal history begins.

Rey's version of swampman allows that general properties might constitute some content for the swampman, but without the specific design history, the specific contents are not fixed; therefore, swamp-people have mental states with indeterminate contents.

Dretske and David Papineau caution against the use of such metaphysical speculation because these are arguably thought experiments that have little relevance to 'facts of life'. The success of the metaphysical speculation also depends on the supposed relation that the representationist argues for between the intentional contents and phenomenal content or character. Is the representationist trying to offer a conceptual analysis of phenomenal content or character in terms of intentional content? If so, mere logical possibilities look problematic. Is the representationist merely offering a 'scientific' identification between phenomenal character and intentional content or a theoretical reduction?[39]

But as Rey points out, surely these intuitions should count for something especially concerning qualitative states (Rey, 1998, p.444). If the teleologist rejects such experiments, then presumably they reject the possibility of the two behavioural indistinguishable individuals - Earthlings and Invertlings - whom possess mental states with distinct intentional contents. But the possibility of such individuals motivates the teleological account of intentional contents as entirely constituting phenomenal content or character,in the first place, as a response to Inverted Earth. Therefore, it is inconsistent of the teleologist to allow for distinct intentional contents in behaviourally indistinguishable individuals, but to then reject swamp-people because they are not metaphysically plausible.

It is always open to the proponent of the teleological theory to 'bite the bullet' and accept the counter-intuitive implication that swamp-people undergo no, or indeterminate phenomenal content or character. But this surely ignores the role of occurrent contents. Accepting the possibility of unfeeling swamp-people does not meet the underlying intuition that

---

[39]See Braddon-Mitchell and Jackson (1997) for attempts to show that such metaphysical speculation is relevant.

grounds the swampman scenario. The evolutionary account cannot take into account the fact that that the swampman has mental states which have very effective current functions that arguably possess their phenomenal content or characterbecause he is in an environment and that environment can enter into the occurrentcontent of that swampman's perceptual states. It is the here and now, specific contents of our mental states that evolutionary accounts just cannot account for. The swampman scenario is just a somewhat vivid scenario that exemplifies this point.

3.4.3 Silicon replacements

David Braddon-Mitchell and Frank Jackson (1997, pp.487-8) reapply a traditional thought experiment whereby a scientist slowly replaces parts of your brain with functionally isomorphic artificial implants against the teleologists. Silicon replacement is arguably not a pleasant prospect, but assuming that the replaced parts fulfil the same current functions, this should be one possible desirable option. Yet according to the phylogenetic or developmentally fixed ontogenetic teleological representationism, this is not a legitimate option, since the artificial parts do not have the right evolutionary history, nor have they been part of the appropriate developmental learning process. The rational response to the artificial replacement is to be concerned only as to whether the artificial parts fulfil the same current functional role as the neurological states before they started to degrade. The *irrational* response would be to worry that the artificial component did not evolve out of natural selection processes, nor was it part of an integrated system that developed and conditioned certain responses to an environment.

> A common thought experiment is to imagine that, as parts of someone's brain degenerate, they are progressively replaced by silicon implants, and to note that, provided the implants fill the pre-degeneration functional roles of the parts they replace, the surgery counts as successful. The key point for us is that what a patient facing such an operation will care about is essentially that what is about to be inserted in her brain will do the job done by the relevant brain states before things started to go wrong (Braddon-Mitchell & Jackson, 1997, p.487).

It might be responded that the silicon parts of the thought experiment are designed, albeit in a way distinct from slow processes of biological natural selection. However, it is not clear as to whether the process of designing of the artificial parts is of any relevance to the concerns of the patient. The only relevant concern to the patient is whether the artificial part has identical current functions to the pre-damaged part it replaces (Braddon-Mitchell and Jackson, 1997, pp.586-7).

Ultimately, however, the above thought experiment only re-enforces the claim that the authors of the metaphysical speculation think that only current functional roles matter. The above argument merely begs the question against the teleologist. It is always an option open to the teleologist to claim that they would not want the replaced parts. Why merely assume that the person who does not care about the causal history of the brain parts is both rational and correct? teleologists might have many independent grounds for maintaining their account so one additional bitter pill to swallow is not so problematic. If one had the option of having a healthy functioning brain part that has the right causal history, perhaps a part grown from your own genetic tissue, or a silicon brain part, it is not immediately obvious that one would choose the silicon implant. But whether *this* choice is just a doubt as to the functional isomorphism of the silicon implant or whether the choice involves a genuine appeal to the relevanceof the design history of the part, is not clear?

However, just as inverted spectrums do not demonstrate the falsity of current long-arm functional Rerpesentationism, the above thought experiments do not demonstrate teleological representationist to be mistaken, but these metaphysical speculations arise out of a deeper underlying assumption that the current functional role of mental states determines the intentional contents of those mental states.[40] In the case of swampman, the current functional roles of his mental states can work

[40] Notice that the possibility of inverted spectrums demonstrates the inconsistency of these two types of representationism. Current long-arm functional representationism cannot allow for the possibility of inverted spectrums across two currently functional isomorphic individuals when they share an environment, but the teleological representationist can allow for differences in two functionally isomorphic, even physiologically identical, individuals, if they have either different ontogenetic or phylogenetic histories.

perfectly adequately; and hence, occurent contents serve to constitute his phenomenal content or character. In the case of silicon replacement, it is assumed that as long as a functioning part does its job adequately, the history is not relevant. It is when the teleologist tries to account for phenomenal content or characterin terms of teleological content that these counter-intuitions become especially strong.

However, the teleologist faces objections greater than the presence of underlying counter-intuitions concerning the relevance of current functional roles. Teleological accounts face substantial problems concerning the vagueness of the intentional contents that these theories are supposed to generate because the teleologist cannot account for specific differences in content if the survival functions come out the same.

3.4.4.1 Indeterminacy of content

Teleological accounts of content are notoriously bad at specifying the determinate contents of mental states. The objections that develop this theme come in a variety of forms. The first identifies an indeterminacy concerningthe referenceclass of that the teleologist claims that the mental states indicate. It is not clear on a teleological account whether frogs detect flies, or small dark moving dots. When applied to Inverted Earth, the survival functions of the Earthling come out the same for the Earthling and the Invertlings. But our phenomenal contents are both specific and highly determinate; therefore, suggesting that a teleological account is not suitable for an account of phenomenal content or character.

The second type of objection identifies a problem with indeterminacy when a perceiving organism moves to a novel environment. This might be applied to Inverted Earth. Why not argue that the Earthling is just moving to a *new* environment? If the Earthling moves to a new environment, why assume that the mental states of the Earthling are malfunctioning? The lenses appear to be affecting the current functions of the Earthling's mental states, but the teleologist does not care about the lenses, given the irrelevanceof current functions.

I do not suggest that there are no adequate replies to some of these objections, but when taken together, they constitute a general problem with adopting a teleological approach to phenomenal content. If there is an alternative solution to Inverted Earth that does not appeal to such teleological 'baggage', I would suggest that this constitutes a superior reply to Inverted Earth.

3.4.4.2 Flies or dots?

The account of content given by the phylogenetic teleologist depends only on the survival functions of the mental states. According to the teleologist the functions of mental states that arise from the mathematics of survival, when spelt out, determines the intentional contents of mental states. Frog's mental states detect flies because 'fly' detecting mental states have a positive fitness value and were selected for in the past to detect flies; therefore, frogs' mental states have the content 'flies'.

> This is where an appeal to teleological considerations seems to yield a natural and satisfying answer. . . we can pick out the truth conditions of a belief as that which it is the biological purpose of the belief to be co-presented with (Papineau, 1990, p.125).

But this answer is not quite so satisfying because the survival functions are the same however the specific contents of the mental states are determined. Distinct contents come out with exactly the same survival value, so it is not clear as to how the teleologist is supposed to determine which specific content constitutes the mental state. Fodor (1994) argues against the phylogenetic teleologist thus:

> [Y]ou can say why snapping is a good thing for frogs to do given their situation, which-ever way you describe what they snap *at*. All that's required for frog snaps to be functional is that they normally succeed in getting the flies into the stomach of the frogs, so long as the little black dots in the frog's normal environment *are* flies, the snaps do this equally well on either account of their intentional objects. The mathematics of survival come out precisely the same either way (p.195).

Applying Fodor's point to Inverted Earth, according to the evolutionary theorist, the Earthling's perceptual system, on Inverted Earth, is working

perfectly well. The perceptual system of the inverted earthling is fulfilling the function it was selected for in the first place, to identify features of an environment and utilise these features, in a manner that will aid his survival. It is absolutely not clear, according to an evolutionary account, as to why representing the sky in some manner is any more relevant to the survival of the Earthling as representing the sky in a different manner as long as all the discriminatory capacities of the perceiving organism are maintained. If inverted spectrums across two individuals fulfil all the required discriminatory functions, then the survival functions all come out the same; therefore, it is not clear as to how an evolutionary account can even begin to offer an account of why mental states were selected to detect, for example, green as opposed to say red.

Papineau (1998) later argues that fly detecting is a mere metaphysical puzzle and should not be considered to invalidate biological investigation.

> Maybe there are underlying metaphysical puzzles about the fact that 'fly' and 'small dark moving thing' are causally relevant descriptions. . . But since nobody would say that these metaphysical puzzles invalidate biological analysis in general, it would be unreasonable to argue that they are an obstacle to a biological analysis of representation in general (p.2).

This is perhaps ironic since in 1990 it was precisely the same metaphysical puzzle that motivated his teleological account. According to Papineau (1990), when a metaphysical puzzle serves to give an opportunity for supporting a theory's plausibility, as a response to the puzzle, the metaphysical puzzle should be embraced. According to Papineau (1998), if a metaphysical puzzle poses a problem to your theory, it should be rejected as a *mere* metaphysical puzzle that has no bearing on 'real' science. Irrespective of Papineau's galling dodge of the argument, the reference class of the frog's mental states might be broader than 'flies' but smaller than 'dots.' Perhaps the mental states of the frog track 'food' (Millikan, 1993). The teleologist at no point commits themselves to such a narrow reference class of 'flies' as the content of frogs' mental states. Alternatively the reference class could be even broader than food. Perhaps frogs' mental states really are selected to detect 'small dark moving objects.' After all it is any member of this class of objects that causes a response in the frog and

it is this mental state that has been selected for to cause a direct response (Neander, 1995). But however the teleologist fixes the referenceclass for what the mental states of frogs represent, it is not clear as to how the phylogenetic teleologist can explain specific colour contents given the identity of survival functions across spectrally inverted individuals.

3.4.4.3 Novel environments

The final objection offered to both phylogenetic and ontogenetic teleological accounts is that they cannot really account for the intentional contents of an organism's mental states changing when it moves into a new environment. According to the teleological accounts the contents of the mental states have been selected for within a specific environment. If a perceiving organism moves into a genuinely new environment, then according to the teleological account, none of the new properties of the novel environment will have been selected for in the history of the ancestors of that organism. None of the new properties of the novel environment will have played any role in the development and conditioning of that organism when it was young. Indeed, it seems essential to the teleological reply to Inverted Earth that the Earthling's new environment has no effect on the intentional contents of the Earthling's mental states in any way. If an evolutionarily selective or developmental history is sufficient for permanently fixing the intentional contents of an individual's mental states, then when this individual moves to a new environment, this individual will not undergo any *new* intentional contents. If phenomenal content or character is entirely constituted by intentional content, then this individual will never see any new objective properties of a new environment. But surely the overwhelming intuition is that when an organism enters a new environment it undergoes a host of new phenomenal contents. Imagine seeing a new shade of purple for the first time, that none of your ancestors have ever been exposed to, or that you, as a child, have never been exposed to.

The best that the teleologist can offer is that if an organism moves to an environment where objective properties change, the intentional contents of that organism is indeterminate. This indeterminacy of content is expressed

by Stich when he considers a bird that flies into a new environment, progressively crossing environments where the proportion of monarchs, an edible butterfly, to viceroys, a poisonous butterfly that very closely resembles monarchs, slowly grows until there are no viceroys at all. The bird presumably develops mental states that indicate both monarchs and viceroys, with the function of avoiding them, as either its ancestors evolved in this environment or it developed this avoidance in its own lifetime. The bird then flies into a new environment where there exist only monarchs. The question is at what point does the bird's mental states cease to represent monarchs and viceroys, and just represent monarchs. At what point does the birds mental states 'malfunction?'

> There is it seems no principled way to answer this question. . . Cases like this are bound to make trouble for Dretske's account of Representation, since if there is no principled way of determining whether C indicates *monarch, or monarch and viceroy,* then there is no principled way of determining what it represents (Stich, 1990, p.805).

Dretske (1990) replies to this argument by pointing out that as soon as a mental state has a non-zero chance of error, it misrepresents. This might have some plausibility when dealing with the contents of beliefs, but is arguably more questionable regarding phenomenal content or character. This is because, arguably, the better explanation of explaining the determinate contents of mental states is the actual properties of objects in the immediate environment of that perceiver and these properties direct and current relation to those mental states of that perceiving individual. Whether the bird perceives a monarch or a butterfly depends on whether it is a monarch or a butterfly that is currently in a direct causal relation to that bird's mental states. If it is this direct relation between the mental state of a perceiver and objective properties of that perceiver's environment that explains the specific intentional content of that perceiver's mental state, then there is no problem in explaining the intentional content of a perceiver's mental states when it moves into a novel environment. Selection histories based on the past evolution of ancestors of a creature and that creature's mental states might tell a worthwhile story about the types of properties a perceptual system detects across a percentage of a population, but when it comes to the specific intentional content of one specific mental state of a

perceptual system, only the actual properties of the immediate environment of that perceiver best accounts for the specific content of that individual's mental states.

On Inverted Earth it appears that our Earthling has moved to a *new* environment, thus there appears to be no real grounds for claiming that his mental states start to malfunction. It might be argued that the Earthlings mental states were selected for in a manner that takes into account the possibility of the Earthling moving to a new environment. The Earthling moves to, not an abnormal environment, but a new environment, thus the Earthling's mental states function perfectly adequately on Inverted Earth. The teleologist just assumes that mental states are selected for in only one 'natural' environment but why assume this? Does an individual's mental states, who was born in Britain, suddenly start to malfunction when this individual travels to America? Surely not. So why assume that the Earthling's mental states start to malfunction when he goes to Inverted Earth. The lenses offer some grounds for maintaining that the Earthling's *current* mental states malfunction, but it is hard to see why the evolutionary history of the Earthling offers any grounds for a malfunction of the Earthling's mental states.

3.4.4 Summing up some objections to teledogical accounts of phenomenal content or character

First, intuition suggests that occurrent contents are to be more closely identified with phenomenal content. It is the current functioning of mental states that establishes a direct causal relation between an object in the environment and a mental state. Metaphysical speculations about swampmen and silicon replacements are only manifestations of this intuition, but arguably do not serve to conclusively demonstrate this intuition. It is not obvious that the teleologist can just reject the scenarios as absurd because it was the metaphysical speculation of Inverted Earth that partly motivates a teleological account of phenomenal content as a response to the argument.

Second, phylogenetic teleological accounts appear to be notoriously bad at specifying determinate contents if two contents result in the same survival value for a creature. This might be illustrated with frogs and flies, but is arguably equally applicable to specific colours given the fact that these might change whilst the survival functions remain constant; thus making it hard for a teleologist to give an explanation of why the mental state that indicates, for example, green was selected. The teleologist cannot deny the possibility of inverted spectrums because it is consistent with their own theory that functionally, even molecularly, identical duplicates can have mental states with distinct intentional contents.

Third, it appears that teleological accounts have problems accounting for how an organism can move into new environments, yet still undergo mental states with determinate contents. This was illustrated with Stich's example of a bird flying into a new environment with different types of butterflies. When considering Inverted Earth, it is hard for the teleologist to identify where the malfunctioning is occurring. It is not just because the environment is new, it is arguably because of the lenses, but the teleologist cannot appeal to the lenses, because the lenses have nothing to do with the historical function of the Earthling's mental states.

Fourth, a problem specific to the phylogenetic account is that one can imagine offspring of the Earthling and Invertlings that inherit the spectrally inverting genetic alteration that was inserted into the Earthling. How many generations have to pass before the teleologist accepts that the descendents have mental states that have an appropriate design function? Any possible answer is surely entirely *ad hoc*; thus suggesting that there is a faulty assumption that the phylogenetic teleologist is making. A further related point is that surely the first creature that undergoes a genetic mutation that offers some major survival advantage, perhaps it enables the creature to perceive colours in the infra-red range, is in its own way a swampcreature. Yet because this was a spontaneous accident, does this rule out this creature as having any content to its infrared detecting mental states?

Because of the above problems with teleological accounts of intentional content as an account of phenomenal content or character, it is suggested

that an alternative representationist reply is preferable as a response to Inverted Earth. There is such an alternative representationist reply to Inverted Earth that makes no appeal to phylogenetic or non-phylogenetic teleological accounts of content.

3.5.1 The real reason Inverted Earth fails to demonstrate the falsity of representationism

To recap, the representationist has two broad options to reply to Inverted Earth; to either: first argue that the phenomenal content or characterof the Earthling's mental states change when he arrives on Inverted Earth and this corresponds to the change in the intentional content of the Earthlings mental states; thus the Earthling's perceptual system overcomes the inverting lenses phenomenally re-inverting (slowly or instantaneously) without the Earthling noticing; therefore, the Inverted Earth argument is unsound. However, an appeal to phenomenal re-inversion was argued to fail in chapter two.

Or second, to argue that the intentional content of the Earthling's mental states stays the same after the move to Inverted Earth, and the intentional content remains different from the Invertlings' intentional contents; thus the phenomenal content and characterof the Earthling's mental states stays the same; therefore, the inverted Earth argument is unsound. We have seen two representationist attempts to argue this line as follows:

First, the narrow content account attempted to show how the narrow intentional contents as functions that map contexts onto contents stay the same after the Earthling arrives on Inverted Earth; narrow contents entirely constitute phenomenal content or character; thus the phenomenal contents and character stay the same; therefore, the Inverted Earth argument is unsound. However, an appeal to narrow contents was demonstrated to fail in chapter two.

Second, the teleological approach argued that the design functions of the Earthling's mental states were fixed either in the phylogenetic or the ontogenetic history of the Earthling; thus the intentional content of the Earthlings mental states were never the same as the Invertlings despite

behavioural identity; thus the intentional contents of the Earthlings mental states did not change because the current long-arm functional roles are irrelevant; therefore, the Inverted Earth argument is unsound. However, there are problems with this account when applied to phenomenal content or character(see earlier)and there is a better reply to Inverted Earth available.

3.5.2.1 The dilemma

The best reply to Inverted Earth exploits an ambiguity in the original formulation of the argument thus allowing the possibility of presenting a dilemma to the proponent of the argument, neither of which options pose a problem to the representationist. The dilemma goes thus: are the inverting lenses that are inserted into the Earthling part of the environment of the Earthling, or are they part of the brain of the Earthling?

3.5.2.2 The lenses are part of the environment of the Earthling

Imagine a different scenario as follows: on Inverted Earth someone paints all the objects the same colours of things as they are on Earth (the sky and the sun are problematic, granted), red grass is painted green, blue bananas are painted yellow, and then an Earthling *without lenses inserted* arrives on Inverted Earth. In this scenario there is absolutely nothing surprising about the intentional contents of the Earthling's mental states staying the same, when the Earthling looks at the same types of coloured objects on Inverted Earth. The phenomenal character of the Earthling's mental states would stay the same when looking at the same types of coloured objects on Inverted Earth. The grass would look just as green and the bananas would look just as yellow. This example is clearly no threat to the representationist.

Now imagine a scenario where objects on Inverted Earth are not painted to resemble the colours of object on Earth; however, in this scenario when Ed arrives on Inverted Earth a bubble surrounds Ed. On the inside of this bubble are projected images of Inverted Earth. The images change as the Earthling moves around and looks at the world creating the illusion that there is no bubble there. These images are, however, altered such that the

colours of the object images are changed so that the same types of coloured object images projected onto the inside of the bubble look just the same as the colour of the same types of coloured objects on Earth. Once again, so far as they concern colours, the contents of Ed's mental states, when looking at the images projected onto the inside of the bubble, are the same as the contents of the corresponding mental states obtained on Earth when looking at the objects corresponding to those images. The bubble is clearly part of the environment of the Earthling and the Earthling clearly and correctly perceives the colour of the images projected onto the inside of the bubble. This example is clearly no threat to the representationist.

Now imagine a scenario where the Earthling wears very light goggles, unbeknownst to him. These goggles have an image of the world projected onto them. Like the bubble, the colours of the images of the same types of objects correspond to the colours of the same types of objects on Earth. Once again, the intentional content of the Earthling's mental states stays the same when looking at the same type of *coloured* surfaces on Inverted Earth. The goggles are part of the environment of the Earthling and the Earthling correctly represents the colours of images on the screen of the goggles. This example is no threat to representationism. Clearly the contents on Inverted Earth differ from the corresponding contents on Earth since the former are about images on either the bubble or the goggles and the latter are about objects beyond the screens, but regarding colour, the differences in content are irrelevant to phenomenal character. So if the lenses of the Earthling are part of the environment of the Earthling, why are the lenses a threat to representationism?

Assuming that the lenses are part of the environment of the Earthling is comparable to the above scenarios. In the case of the colour inverting lenses, there are no painted objects or painted images; however, if the lenses are part of the environment, the Earthling correctly represents colours. The mental states of the Earthling still detect the right wavelengths of light entering the Earthling's eye. There is clearly a sense in which the Earthling is fooled concerning the actual colours of, for example, ripe tomatoes, but this fooling is no more mysterious than when someone paints the Inverted

Earth objects to look like Earth objects, or altered images are projected onto the inside of a bubble.

In the case of the bubble, the goggles or the lenses, the Earthling and the Invertlings do not share an environment. In the environment of only the Earthling there is interposed between the Earthling and his environment some mechanism, which is part of the environment, specific to the Earthling. In one sense the Earthling and the Invertlings share an environment; i.e. they are on the same planet, but in a crucial sense they do not share an environment because the Earthling has as part of his environment a further mechanism that the Invertlings do not. The Earthling with lenses inserted is, to all extents and purposes, part of a goggled or bubbled world while the Invertlings are not.

Because the Invertlings and the Earthling do not share an environment, this immediately undermines any attempt to appeal to Invertlings to justify a sameness of intentional content between the Invertlings and the Earthling. This thus undermines an attempt to appeal to the intentional content of the Invertling's mental states to justify a change in intentional content in the mental states of the Earthling, when the Earthling moves to Inverted Earth. Furthermore, if the Invertlings are not appealed to, seeing as the Earthling is representing colours perfectly correctly when presented with same types of coloured objects, there is no reason to suppose that there is a change in the intentional content of the Earthlings mental states when he arrives on Inverted Earth. There is, therefore, no change in the Earthling's phenomenal content or character, when presented with the same patterns of light. *If the lenses are part of the environment of the earthling then there is no need to maintain that the intentional content of the Earthling's mental states changes; therefore, the Inverted Earth argument is no threat to the wide representationist*

3.5.2.3 The lenses are part of the brain of the Earthling

If it is assumed that not the lenses, but a tampering of the brain of the Earthling brings about the colour inversion, it is no longer clear that there is any sense in which the Earthling correctly represents colour. If the

tampering is part of the brain, there is no reason to suppose that the phenomenal content or character of the Earthling's mental state stays the same when the Earthling simultaneously looks at the same types of *objects* on Earth and Inverted Earth. There is a sense in which the Earthling notices no difference when he wakes up on Inverted Earth, but athough he does not notice that anything phenomenal has changed, a phenomenal change can easily be demonstrated to him by presenting the Earthling simultaneously with an Earth object and the same type of Inverted Earth object.

If he was to simultaneously look at a ripe Earth tomato and a ripe Inverted Earth tomato on Inverted Earth to compare the look of the same types of object, the ripe Earth tomato would look green to the Earthling because of the brain tampering. The ripe Inverted Earth tomato would look red to the Earthling. Therefore, the same types of objects if presented simultaneously, *do not look the same* to the Earthling. If the brain of the Earthling is altered, then the look of the same types of objects on Earth and Inverted Earth do not stay the same if the Earthling was to be simultaneously presented with the same Earth and Inverted Earth objects after the Earthling moves to Inverted Earth It therefore cannot be inferred from the Inverted Earth argument that the phenomenal content or character of the Earthling's mental states stays the same if the Earthling is presented simultaneously with an Earth object and the same type of Inverted Earth object.

There is an essential sense in which the phenomenal content or character of the Earthling's mental states changes if the Earthling were to look at the same types of Earth objects on Inverted Earth. It is this ambiguity that partly motivates the Inverted Earth argument. If the lenses are part of the Earthling such that he really does fail to detect wavelengths correctly, there is no sensible sense that can be made of the idea that the Earthling represents colours correctly on Inverted Earth or that there is a constancy in phenomenal content or character when the Earthling looks at objects of the same type on Inverted Earth when compared to objects of the same type on Earth. Furthermore, his mental states will never come to represent colours correctly either on Earth or Inverted Earth. Given this change in the phenomenal content or character corresponding with the intentional contents of the Earthling's mental states mis-representing, there is no reason to

maintain that phenomenal content or character is distinct from intentional content. The proponent of the Inverted Earth argument requires a constancy of phenomenal content or character across the simultaneous inspection of Earth and Inverted Earth objects, and this is not established by the Inverted Earth argument.

If the lenses are built into the brain of the Earthling, there is no reason to suppose that there is any phenomenal constancy, if the Earthling were to look at the same types of objects, before and after the move to Inverted Earth; therefore, the Inverted Earth argument fails to establish that phenomenal content or character is entirely constituted by intentional content.

3.5.3 But what about functional isomorphism, causal co-variation and linguistic embedding?

Imagine that our Earthling returns to Earth with an Invertling. The Earthling is most surprised and notices how odd everything looks. He describes tomatoes as green, etc. The Invertling is in complete agreement with Ed. He also points out how odd everything looks. However, from the fact that both the Earthling and the Invertling are in complete linguistic agreement, functionally isomorphic, and even causally co-variant with the same types of objects, and the same types of coloured objects, on Inverted Earth, it absolutely does not follow that the Earthling correctly represents the colour of objects. The Invertling represents colours correctly. The Earthling misrepresents because of the tampering with his visual system. The Earthling misrepresents not because of teleology. The intentional contents of the Earthlings mental states are clearly distinct when compared with the Invertling. The phenomenal content or characterof the Earthlings mental states change after the tampering with his visual system because it is now clear that the same types of objects do not look the same way as they used to look before he had his perceptual system tampered with. He can now compare the looks of the same types of objects, for example, ripe Earth tomatoes and ripe Inverted Earth tomatoes. The only sense in which the phenomenal content or characterof the Earthling's mental states stays the same, assuming that the tampering is part of the Earthlings processing, is

when the perceptual experiences of the Earthling looking at Inverted Earth objects is compared to the Earthling looking at the same coloured types of objects.

If the lenses are part of the Earthling, his mental states fail to ever represent colours correctly on Earth or Inverted Earth, and hence there is good reason to maintain that both the intentional content and the phenomenal content or character of the Earthling's mental states change if he were to simultaneously look at the same types of objects on Inverted Earth and on Earth. The Earthling's mental states permanently mis-represent the colours of objects. If this horn of the dilemma is accepted, the phenomenal content or characterof the Earthling is permanently tampered with and the Inverted Earth argument is unsound.

3.5.4 The successful representationist reply concluded

If either the lenses are part of the environment of the Earthling, or built into the neurological structure of the Earthling, then the Inverted Earth argument does not succeed in establishing that representationism is false. If the lenses are part of the environment of the Earthling, this is analogous to placing the Earthling in a painted, bubbled or goggled world. This presents no problem to the representationist. The Earthling has no problem correctly representing the same types of coloured objects. There is no reason to maintain that the intentional contents of the Earthlings mental states or the phenomenal content or character of the Earthling's mental states changes, *but not on grounds of phylogenetic teleology or developmentally fixed ontogenetic teleology.* If the lenses are integral to the Earthling, then the Earthling mis-represents intentionally and phenomenally the same types of objects on Earth, Inverted Earth and everywhere. The Earthling permanently mis-represents the colours of objects. There are no grounds for maintaining a phenomenal re-inversion if the Earthling is placed on Inverted Earth; therefore, if the lenses are part of the Earthling, then there are no grounds for maintaining that the phenomenal content or characterof the Earthling *stays the same when the Earthling simultaneously looks at the same types of Earth and InvertedEarth objects* on Inverted Earth. Either way there are no good grounds offered by Inverted Earth for driving a wedge between

intentional content and phenomenal character. The Inverted Earth argument fails to establish the falsity of representationism.[41]

3.6 Inverted spectrums again

This reply to Inverted Earth also suggests a plausible representationist reply to intra-personal spectrum inversions. Phenomenal re-inversions cannot be ruled out in the four stage intra-personal inversions because there is no requirement for an unnoticed phenomenal re-inversion, and there is possibly a presence of behavioural readjustment in the four stage intrapersonal inversion scenario that is lacking in the Inverted Earth scenario. However, the above dilemma also applies to the intra personal inversion scenario; are the lenses part of the environment or part of the Earthling?

If the lenses are part of the environment, then there is a sense in which the individual who has the lenses inserted does not phenomenally or intentionally change at stage two, when the lenses are initially inserted. Therefore, the individual never comes to intentionally revert back to the way he once was. Just think of the lenses as analogous to painting every object a different colour. The individual may change his behaviour, but the colours of objects are permanently changed and he sees these changed colours *correctly.* There is a sense in which the same objects look a different colour but this is no more a threat to the representationist than painting ripe tomatoes green. This individual cannot be compared to other individuals in his community to justify a possible *change* in the intentional content of the mental states of the individual; therefore, an essential motivation for thinking that the intentional contents of the individual changes is undermined.

If the lenses are part of the individual, then there is a clearer sense in which the phenomenal content and character of his mental states changes. His perceptual systems have been tampered with in a way far more fundamental

---

[41] Notice also that this reply makes no appeal to narrow content accounts to establish that the intentional contents of the Earthling stay the same. The reason that the intentional contents might be argued to stay the same stay the same is because the lenses might be assumed to be part of the environment of the Earthling.

than painting objects a different colour. He no longer perceives colours correctly, nor will he ever perceive colours correctly. However long he remains phenomenally re-inverted, he will always remain phenomenally re-inverted. His intentional contents will permanently change corresponding to the phenomenal change. Notice that this is despite linguistic agreement and coarse functional isomorphism with all his companions.

3.7 Conclusion

The first part of this thesis considered two metaphysical speculations as an attempt to motivate phenomenism out of presenting objections to representationism. Both metaphysical speculations failed to demonstrate the falsity of representationism. However, both arguments served an invaluable role in motivating both different types of representationism as responses to these metaphysical speculations and apparent inconsistencies with certain existing representationist responses to these thought experiments. It is perhaps revealing that the teleological responses only relied on the historical design functions of mental states in an attempt to maintain intentional constancy in the Earthling while if they looked more closely at the failure of the current functioning of the Earthling's mental states, they would have realised that the Earthling's mental states badly misrepresent the same types of coloured objects. This was simply shown by returning the Earthling to Earth. Narrow content replies are motivated out of making the Inverted Earth experiment work, but given its failure, this offers a further undermining of narrow content accounts. The serious misrepresenting of the Earthling's mental states might be hard to reconcile with the supposed success of the narrow contents. Furthermore, as established in the last chapter, narrow content accounts of intentional content face an insurmountable problem. Teleological accounts of intentional content are too clumsy to specify the determinate contents of individual phenomenal states. This leaves the current long-arm functional representationists or one version of wide representationism.

The wide representationist, however, faces problems entirely independent of the considerations outlined in the first part of this thesis. Wide

representationists face deep problems concerning the casual efficacy of relational properties. These considerations have been argued, to and fro, with regards to mental states such as beliefs, but these considerations are equally applicable to the representationist who claims that phenomenal character is entirely exhausted by wide intentional content. If the wide representationist wants to give phenomenal character some causal or deep explanatory role, then these same concerns apply also to them. Part two of the thesis examines the casual or explanatory implications for wide representationism and phenomenism.

Part II
Explanation and Causation

Chapter 4: Explanation, Causation and Wide Content

4.1.1 Explanatory constraints and wide content

In the first part of the thesis, two types of metaphysical speculation were considered as an attempt to address possible motivations for phenomenism by presenting arguments against representationism. These arguments against the representationist were concluded to fail. Interpersonal spectrum inversions beg the question against the representationist, intra-personal spectum inversions allow for the possibility of phenomenal re-inversion, and there is an ambiguity being exploited by the phenomenist concerning the status of the lenses as either being part of the environment or integral to the perceptual processing of the individual wearing the lenses, this latter point is also exploited by the phenomenist in the Inverted Earth argument. Once this ambiguity is clarified, both intra-personal and Inverted Earth arguments can be demonstrated to be unsound. However, a narrow content account of phenomenal character, inspired by a Fodorian account of narrow content, was found wanting as possible intentional identification of intentional content with phenomenal content or character. Suspicion was also cast upon historical or teleological accounts of intentional content when applied to phenomenal content or character. This leaves only a version of wide representationism as an adequate representationist account of phenomenal content or character.

Part two of the thesis develops some more considerations that might be thought to count for or against wide representationism and phenomenism; these considerations, although sometimes assumed to rely only on metaphysical speculation are perhaps grounded instead by matters of fact. The considerations that concern part two are considerations that have traditionally challenged both phenomenism and wide intentionalist theories of mind, although lately wide intentional theories of mind have received the

brunt of these objections mainly due to Fodor (1987). These considerations concern the causal or explanatory role of mental states.

Constraints on causation and explanation present a troubling challenge to the wide representationist. Furthermore, phenomenism as a theory of phenomenal character *appears* better suited to handle what might be described as 'causal or explanatory principles', although whether phenomenism is better able to handle these causal or explanatory considerations will be considered in chapter six. This chapter elucidates some of these causal or explanatory principles, examines how they might be taken to be inconsistent with wide representationism, and analyses the success of two representationist replies to these inconsistencies: Wilson (1997) and Yablo (1997). The onus is placed on the wide representationist to first meet these principles of causation or explanation, and second, present an adequate theory of causation or explanation that accommodates wide representationist intuitions. If the causal-explanatory challenge is presented adequately, it appears that there remains a large lacuna in this representationist theory, when it comes to their attempt to explain how wide contents can be causal or explanatorily efficacious. If the wide representationist wishes phenomenal content or character to have some causal or explanatory role, then the general objections that apply to wide intentionalist theories of mind also apply to the wide representationist who maintains that wide intentional content entirely constitutes phenomenal content or character.

Through the development of this argument, some curious dynamics arise between types of reasoning within the philosophy of mind concerning the metaphysical and logical analyses of relational properties and constraints within scientific explanation. The reason for concluding that considerations about causation and explanation count against the wide intentionalist is summarised by Crane (1991); he applies the argument to wide intentionalist theories of beliefs and desires:

> The problem . . . arises from [the] conflict with two fundamental principles about causation and the nature of intentional states. The first principle is that intentional states have causes and effects. I take it as uncontroversial that beliefs, for example, are caused by perceptions and other beliefs, and combinations of beliefs and desires

> cause actions. . . . The second principle is about which properties of things are involved in causal interactions between them. . . . We should distinguish between those properties whose acquisition or loss by a particular is a *real change* in that particular and those whose acquisition or loss is not one. . . Properties of the first kind are, uncontroversially, *intrinsic* [non-relational]properties; and properties of the second kind are non-intrinsic [relational].
>
> [W]hen we look for causes, we look for intrinsic [non-relational] properties. . . . [B]road intentional states are not, and do not systematically depend on, intrinsic properties of thinkers . . . [T]herefore wide representationists] must deny either that intentional states as such have causes and effects, or that relational properties . . . cannot have immediate effects (pp.5-9).

I will also take it as uncontroversial that intentional mental states, including phenomenal mental states, both cause and explain behaviour. Fodor appeals to scientific practise as part of his formulation of his argument (Fodor, 1987). Fodor appeals to scientific practise to justify why relational properties are not causally efficacious. However, Crane (1991), when formulating this argument, claims that an appeal to scientific practise as part of the argument would be a mistake.

> Fodor weakens his case by making the dependence on intrinsic properties a high-level generalization of scientific practice. [However, these two principles] are not just generalizations about the current practice of science, but assumptions about causation and the causal status of mind (p.8).

I disagree with Crane. The anti-wide representationist owes the wide-representationist some explanation as to why relations cannot cause. It is precisely the causal efficacy of relations that seems, in the current literature, to be sometimes defended (Wilson, 1997, Yablo, 1997). Crane's assumption about relational properties and causation cannot be assumed; this assumption requires justification. This justification for the causal or explanatory problems that confront relational properties arises from causal and explanatory principles. Therefore the anti-wide representationist can considerably clarify his argument by a development of why relational properties are not suitable for playing a deep explanatory role by an appeal to causal or explanatory principles. There are a host of possible confusions that arise from an analysis of relational properties and their causal efficacy,

so before an investigation of relations in terms of scientific principles is developed, some preliminary points have to be addressed.

First, there is a variety of possible explananda that the wide representationist might be taken to be attempting by an appeal to relational properties. The wide representationist might be attempting to explain variables associated with one of the terms of the relation by variables associated with the other term of the relation. The wide representationist might be attempting to explain variables associated with a third event, entirely distinct from the terms that constitute the relation, by the relation. The wide representationist might be attempting to explain this third event by an appeal to variables associated with only one of the terms of the relation.

Second, there has to be some clarification of the wide representationist's commitment to relations as explanatory. Sadly, the nature of the relations to which the wide representationist is committed is rarely spelt out. This chapter will present two conceptions of relations: a *strict* conception, whereby X stands in a strict relation to Y if it is possible that this relation should cease to hold solely on account of a change in Y (this corresponds to Crane's distinction between intrinsic and non-intrinsic properties);[42] and a *looser* conception, whereby there are no constraints on the changes in variables associated with the two terms that constitute the relation. Wilson (1997) offers an explicit rejection of strict relations as causally efficacious in an attempt to justify wide causation:

> Many relational properties, however, *are* causally efficacious in a way that the paradigmatic cases of mere-Cambridge properties are not. For example, being a mother, being unemployed, being a member of a particular species, being a planet, being located in a magnetic field and occupying a relatively specific ecological niche are all relational properties that different entities can have in particular instantiating circumstances, each of which, when coinstantiated with the appropriate properties, enables an entity to bring about particular effects (p.124).

---

[42]Otherwise known as mere-Cambridge properties (more on these later).

Therefore, it might be thought that explanatory constraints only present a challenge to the strict conception of relations. However, the next chapter argues that these explanatory constraints not only present a challenge to the strict conception of relations, but also to the looser conception of relations. Attempts to justify the explanatory relevance of the looser relations are problematic since the wide representationist still has to go beyond the loose relation to demonstrate how the variables associated with the terms that constitute the loose relation meet explanatory constraints. The wide representationist's attempt to justify the explanatory relevance of the loose relations inevitably entails that the *relation* fails to deeply explain in a manner that provides an understanding of the event or process.

Finally, the challenge to the causal efficacy of strict relations arises not from one, generalized, high-level, *a priori* causal constraint, but a myriad of methodological constraints placed on explanations in a variety of different explanatory contexts. These constraints push the wide representationist to the edge of legitimate explanation, and arguably over the edge.

Before, the argument against the wide representationist is developed it is worth clarifying further as to whether the argument is about whether mental states should be taxonomised by their wide intentional contents. Does the argument apply only to strict relations or loose relations? Does the argument rely on one explanatory constraint or many? Are these explanatory constraints *a priori* or *a posteriori*?

4.1.2 Taxonomies, superficial explanations or deep explanations

Before explanatory constraints are considered and applied to relational properties it should first be clarified that the concern of this chapter is not the traditional debate between Tyler Burge (1986) and Jerry Fodor (1987) concerning how mental states should be taxonomised or individuated. It might be argued that the explanatory relevance of the relational properties that determine wide contents does have a direct bearing on how mental states should be individuated (Fodor, 1987, pp33-34). It might even be argued that the way mental states are taxonomised has a direct bearing on the deep explanatory role of such states (Wilson, 1997, pp.124-125).

I want to attempt to avoid this debate and focus on whether relational properties, which determine the wide contents of mental states, have a deep explanatory role to play and one of the terms of the relation is either the neurological or intrinsic states of an individual when the third term, independent of the relation and to be explained by the relation, is the ensuing behaviour of that individual. As stated earlier, I assume that the representationist who claims that wide contents entirely constitute phenomenal content or character assumes a deep explanatory role for phenomenal content or character concerning behaviour. In the next chapter it will also be considered as to whether a looser conception of relations is suitable for an explanatory role for the intentional contents that the wide representationist maintains entirely constitutes phenomenal character or content.

4.1.3 Relational properties

As these various constraints on explanation are offered, they are going to be presented as a challenge to a specific set of properties, relational properties. It is a feature of wide representationist accounts that they are often somewhat under-specific about which terms constitute the relevant relations; however, when they maintain that *wide* contents explain, it is arguably certain that at least one of the terms of the relation resides at some 'distance' from the other term. Sometimes critics of the anti-wide representationist argument assume that the anti-wide representationist holds all relational properties to be entirely redundant with regards to *all* causal efficacy and explanation. This is not necessarily the case.

Various confusions might arise concerning how relations should be analysed. Let us start with a simple relation aRb, where a and b are terms or events with various variables associated with these events and R is a relation. For example, Bob is taller than Joe. Bob is 'a' and Joe is 'b' and the relation is 'taller than'. The specific variable associated with Bob might be him possessing the property of being 6 feet tall; the specific variable associated with Joe might be him possessing the property of being 5 feet tall. One confusion arises concerning what explains what; whether the

terms that constitute the relation themselves are explanatory with respect to each other, or whether the *relation* itself is causally explanatory either in explaining the variables associated with the terms of the relation, or whether the *relation* itself is explanatory with regards to an additional third event 'c'.

Depending on how the explanatory role of relations is construed, the anti-wide representationist need not maintain that all terms that constitute any relations are explanatorily redundant with regards to one another. For example, the atomic explosion occurred 'before' everyone within a certain radius of the explosion died. These two events are clearly causally related. Just because a relation, for example 'earlier than', links two terms does not necessarily entail that there need be no causal relation between variables associated with these two terms, although relations *might* be defined in such a manner that rules out an intrinsic change in at least one of the terms that constitutes the relation. Therefore, the anti-wide representationist further needs to argue that the type of relation that the wide representationist appeals to are only explanatorily redundant with regards to the events or variables that the wide representationist hopes these relations explain. One term of the relation to which the wide representationist appeals is the intrinsic state of the agent. The third independent term that the wide representationist argues is explained by the relation will be designated as the ensuing behaviour of the agent.

If the explanatory emphasis is the relation, and not on the terms, then the anti-wide representationist might, however, plausibly maintain that *relations,* themselves, never causally explain, perhaps because strict relations are not *bona fide* events. Strict relations are not events because strict relations alone cannot change.

> Now all change consists in some object's having some property and then lacking some property; but a relational change consists in an object's having and then lacking a property solely in virtue of the fact that object bears some relation to an object which 'really' changed or altered. Xanthippe's becoming a widow consisted entirely of her being the last wife of Socrates when he died (Lombard, 1999, p.364).

Brian Lombard has a strict conception as to what constitutes a legitimate relational change. He illustrates this with an example of temperature applied to distinct objects and the relation 'cooler than' applied to the two objects. He considers two distinct scenarios as follows: The first is whereby one object is 40 degrees and the other is 50, and then at a later time the first object is *heated* to 60 and the second is *cooled* to 45 degrees. The second object stands in the changing relation 'being cooled relative to the first object'. Lombard describes this *apparent* change in the relational properties of the second object as non-relational because there is a genuine change in the temperatures of the second object; it really does become cooler as the first object becomes hotter; therefore, what appears to be a relational change is really a non-canonical description of an event.

In the second scenario, the first object is heated from 40 to 60 degrees, but the second is maintained at 50 degrees. The second object still remains in the 'being cooled relative to the first object' relation. However, there is no change in the properties of the second object other than the relational property; therefore, this constitutes a genuine relation (Lombard, 1999, p.365).

Even if one allows a dual role for the relation of being cooled relative to another object, which allows for a change in variables associated with both terms of a relation, relations are dependent only on changes in variables associated with the terms that constitute the relation. Lombard draws a rather stark contrast between relational and non-relational properties whereby it is the *non-changing* status of variables associated with one term relative to *changes* in variables associated with the other term that defines a relational property. I would be inclined to claim that both examples of relational properties are legitimate examples of the type of relation the wide representationist might appeal; however, even assuming the looser conception of relations, the changes in the relational property is *apparent* and only occurs because of changes in the variables associated with the terms that constitute the relation.

Relational properties might be employed by the representationist to try to explain variables associated with a distinct event. For example, the

relational property of one person being taller than another appears to explain why one person and not the other reaches the cookie jar without standing on a chair. The problem confronting the wide representationist arises concerning either whether relations themselves are explanatorily relevant, or whether it is the variables associated with the terms that constitute the relation and the underlying causal mechanisms, that are explanatorily relevant relative to a third event, bearing in mind that one of the terms, according to the wide representationist is a state of an agent and the third event to be explained by the relation, is the behaviour of that agent. If one person is taller than another, this is of no relevance to explaining the specific heights of the two people. If the taller person grows a couple of inches and the other person stays the same height, then the relation itself appears to change; i.e., one person gets *even* taller than the other; however, this does not in any way affect the height of the person who remains the same height. If the taller person grows 2 inches and the shorter person shrinks by 1 inch, then the taller person grows *even* taller, relative to the shorter person. Furthermore, there is a change in the variables associated with both of the terms that constitute the relation, but does the apparently changing relation explain the change in the variables associated with the two terms? Does the apparently changing relation itself explain any third event such as the abilities of individuals to reach objects at a certain distance? Answers to these questions will be developed in this and the next chapter.

The anti-wide representationist, in order to demonstrate a problem with wide representationism, has to establish the following: first, if the relation is strict, then the terms that constitute the relation that the wide representationist hopes to be explanatory, have no explanatory relevance to one another, or to a third distinct event, one of these terms being the intrinsic state of a person, and the third event being the behaviour of the person; or second, if the relation is loose, then *relations themselves* are of no causal relevance to changes in the terms that constitute the relation; or the *relation* between two terms is of no *deep* explanatory relevance to the distinct event that is hoped by the representationist to be explained by the relation.

4.1.4 *A priori* or *a posteriori?*

The anti-wide representationist demonstrates that the variables associated with the terms that constitute relations are not of explanatory relevance to one another, or a third event, by an appeal to scientific practice; therefore, the anti-wide representationist has good grounds for attempting to place constraints on explanation. However, the wide representationist often finds it easy to reply to these constraints because the anti-wide representationist often fails to present these constraints as *methodological*, thus allowing the wide representationists to rebut by scoffing at over-generalised, individual, illegitimate, *a priori* attempts to place limitations on scientific endeavour.

> Sciences, particularly the special sciences, develop their own taxonomies, theories and explanations in response to very different research interests and problems, and with very different methodologies. Because of the often unappreciated, rich diversity within the sciences, any argument that relies on a premise about scientific taxonomies and explanations having some essential feature is unlikely to be sound; more so any argument that relies on a general premise about the notions of causation, property or explanation, notions at least some of whose principal instances are found by turning to the sciences. The pragmatics of scientific explanation cannot be separated from its metaphysics; at least not if the metaphysics one proposes for science is to be taken seriously as an account of the metaphysics science actually traffics in. Scientific practice is not simply the ultimate arbiter for claims about the nature of science; its examination is the way to do philosophy of science. Although this conclusion may well strike one as trite, we do well here to remember that philosophers, including philosophers of science, tend to crave the general. Insofar as this is a craving satisfied only by a priori argumentation, it should be resisted (Wilson, 1997, p.138).

Wilson is perhaps correct in expressing a concern that any one individual constraint on explanation offers a constraint on *all* scientific endeavour, but he is perhaps wrong in assuming that many constraints on a variety of scientific endeavours present *no* challenge to the wide representationist. What if all constraints within many explanatory contexts, including the special sciences, prove troubling to the wide representationist? Wilson is perhaps correct in expressing a concern about *a priori* constraints on scientific explanation. However, it is not always clear that such constraints are entirely motivated by *a priori* reasoning, but instead are motivated by an examination of scientific practice and the underlying assumptions that

investigators make to rule out certain lines of inquiry before the investigators begin to investigate. Perhaps Wilson is once again wrong in assuming that the pragmatic constraints on scientific endeavour, within a variety of specific explanatory contexts, present *no* challenge to the wide representationist.

Therefore, to increase the plausibility of the causal or explanatory challenge against the wide representationist's appeal to relations, various pragmatic constraints on causation and explanation will be presented. Some care is going to be taken to show how multiple causal or explanatory constraints present a challenge to the wide representationist who maintains that wide content, which is constituted by relations, and only wide content entirely constitutes phenomenal character. Thus four constraints will be considered that emerge out of various theories of causation and explanation as follows: a non-causal nomological account of explanation, a probabilistic causal theory, a counterfactual theory of causation, and a causal process theory.

The first explanatory constraint might be motivated by Russell's comments, largely inspired by Hume, regarding the role of explanation in science. According to this account, the concern of science is not causal efficacy, but *a posteriori* investigation and the identification of functional relations between variables. Science, through the process of empirical observation, identifies correlations and regularities rather than necessary causes. Thus methodological constraints on explanation need not appeal to *a priori* causal constraints; for example, a non-causal nomological law account of explanation also suggests a possible constraint that might rule out certain types of events as explanatorily useful. Such an explanatory constraint, based on functional correlations between variables, poses the *first* challenge for the wide content theorists appeal to relations.

4.2.1.1 Nomological laws and content

The general purpose of physics as a scientific enterprise is not often to explain the occurrence of individual events, although individual events may end up being explained. The emphasis is usually on the effects on the variations of variables and establishing functional relations between

variables. Some common questions in physics are: What happens if we keep this variable constant whilst we change another variable? If we increase one variable, does the other increase? If we decrease one variable, does the other decrease? By how much? Once correlations are discovered and mathematical functions between variables made explicit, individual phenomenon may be predicted and explained in terms of general nomological laws of nature. " [T]he question '*Why* does the phenomenon happen?' is construed as meaning 'according to what general laws, and by virtue of what antecedent conditions does the phenomenon occur?'" (Hempel & Oppenheim, 1965, p.58).

The first suggested pragmatic constraint on explanation is that for one variable to be of any explanatory relevance to another variable, there must be some *specific* functional correlation relating the variables. This functional correlation usually adopts the form of an equation relating the variables and is sometimes described as a nomological law of nature; for example, Boyle's gas law relates pressure, volume and temperature: PV = nRT.[43] From this law, looser generalisations can be presented in a conditional form, such as, if the temperature of a certain fixed volume of gas is increased, then the pressure within that volume also increases. Specific instances may then be explained in terms of this looser, conditional generalisation or in terms of the specific law; however, according to this Hempelian deductive nomological model, there must be some specific functional correlations, between the relevant variables from which the looser generalisation is derived, for these variables to be of any explanatory relevance to one another.

Consider billiard balls: Billiard ball 'a', with a certain mass, on a very smooth billiard table, moves at a certain velocity towards stationary billiard ball 'b'. Billiard ball 'a' hits billiard ball 'b'. Billiard ball 'b' starts rolling towards a pocket at a certain velocity, eventually falling into the pocket. Billiard ball 'a' carries on moving with a certain velocity. Why does billiard ball 'b' fall in the pocket? A simple law, incorporating a principle of the conservation of energy, relating masses and velocities, combined with

---

[43] The fact this law is only approximate and is thus strictly false is irrelevant; it is merely cited for illustrative purposes.

a set of initial antecedent conditions, explains why billiard ball 'b' falls into the pocket. An anti-wide representationist argument, drawing on this explanatory model, can be run against the wide representationist.

The Hempelian deductive nomological model suggests the first premise. According to the Hempelian model of explanation, a functional correlation between variables is necessary in order for one variable to be explained in terms of the other. The second premise of this version of the anti-wide representationist argument is that the variables associated with the term in one part of the relation that determines the wide contents never functionally correlates with variables associated with the second term, the intrinsic mental states of a person. Furthermore, the relation itself could never functionally correlate with variables associated with a third event, the ensuing behaviour of that person. Thus, no laws of nature exist among the variables associated with the events the wide representationist hopes are explanatory with regards to each other.[44]

It is therefore the relational nature of wide contents that presents a problem in identifying functional correlations between wide contents and certain events. Fodor (1987) gives an account of a relational property that he takes to be analogous to the kind of relational property to which the wide representationist putatively appeals. He presents an example whereby the state of every particle in the universe depends on whether one turns a coin heads or tails up. The two terms of the type of relation to which a wide representationist putatively appeals are as follows: the first is the orientation of a dime that is either heads or tails up; the second term of the relation is the state of every particle in the universe. If the dime is heads up, then every particle stands in a 'heads up' relation to the dime and might be described as 'H' particles. If the dime is tails up, then every particle stands in a 'tails up' relation to the dime and might be described as 'T' particles. More formally, aRb, whereby 'a' is the dime and 'b' is any particle in the universe. R is a relation dependent on the state of 'a'. If 'a' is H (heads up) then R is a 'heads up' relation and 'b' is described as a 'H' particle. If 'a' is T (tails up), then R is a 'tails up' relation and 'b' is described as a 'T'

---

[44] Davidson also argues for the anomalism of the mental, but on distinct grounds.

particle. Notice that this relation is consistent with Lombard's strict criteria for a relational property because all variables, apart from the relational property, associated with the second term remain entirely unchanged when only variables associated with the first term change. The relational property associated with the second term changes when there is *only* a change in variables associated with the first term.

The problem is that there are no functional correlations required between variables associated with the two terms that form the relation, in this case the orientation of the dime and variables associated with, for example, an electron on Alpha Centuri. Turning the dime over in Fodor's pocket will not make this electron spin faster or jump an energy level. Strict relational properties fix the variables associated with one of the terms of the relation. This necessarily entails no functional correlation between variables associated with the terms of the relation.

> How on Earth could the causal powers of particles on Alpha Centuri depend on the orientation of my dime? Either there would have to be a causal mechanism to mediate this dependency, or it would have to be mediated by a fundamental law of nature; and there aren't any such laws. Of course there aren't (Fodor, 1987, p.39).

If the wide representationist relies on relational properties analogous to, for example, the orientation of a dime and the electron on Alpha Centuri, then there need be no functional correlations between the terms that constitute the relation, nor is it clear how there possibly could be any functional correlation between these variables. For a strict conception of relations, it is necessary that the variables associated with the second term do not change, other than with respects to the relational property, and this change is only because of a change in variables associated with the first term. If the variables associated with the second term necessarily do not change, how can there be any functional correlations between variables associated with the first term and variables associated with the second term? How can there be any functional correlations between un-changing variables associated with

the second term and a third event that the wide representationist hopes to explain?

Consider, once again, the relational property: 'being taller'. x is 5 feet tall and y is 6 feet tall. y is taller than x. Is this kind of relational property subsumable under some functional correlations between distinct events? Certainly this relational property *seems* to have explanatory relevance with regards to a distinct event, for example, explaining why y can reach the cookie jar without standing on something or jumping, and x cannot. However, it is arguably not the *relation* that explains why y, and not x, can reach the cookie jar without standing on a chair or jumping. It is the specific height of y, combined with his reach, which explains why he can reach the jar. It is the specific height of x, combined with his reach, which explains why he cannot reach the jar. The relation explains nothing, although it does serve to quantify one variable, i.e., height, as greater, between the two individuals.

One might construct a generalisation thus: in order for p to reach q (an object above p), without p standing on something or jumping, the distance of q from the ground must be no greater than the distance between the toe and finger tip of p, with p's arms stretched to the maximum in a straight line. The distance from Bob's toe to his fingertip with arm outstretched in a straight line is 7 feet. The distance of the cookie jar from the floor is 6 feet. Therefore, this explains why Bob can reach the cookie jar, without standing on something or jumping. The distance from Joe's toe to his fingertip with arm outstretched in a straight line is 5.5 feet. This explains why Joe cannot reach the cookie jar without standing on something or jumping.

If the wide representationist maintains that the relational properties that determine wide contents are of a type analogous to the orientation of a dime relative to the causal powers of an electron on Alpha Centuri, then it appears that Fodor has a point. Wide representationism maintains that relational properties of a certain type determine the content, and hence, relational properties determine the phenomenal character of mental states. According to wide representationism, if the relational properties of two mental states

are distinct, it then follows that those mental states are distinct. If the relevant relational properties of a phenomenal state are distinct, then, according to the wide representationist, the phenomenal characters of these two mental states are distinct. However, variables associated with relational properties require no functional correlation to variables associated with relevant mental states and ensuing behaviour, just like the orientation of a coin relative to the state of particles on Alpha Centuri. There are no nomological laws of nature that correlate the variables that determine contents with mental states and ensuing behaviour. According to the above constraint on explanation, if there is no law of nature, i.e., no functional correlations between variables, then these variables have no explanatory value. Therefore, wide content, and hence phenomenal character has no explanatory value with regards to the ensuing behaviour of an individual.

4.2.1.2 Laws again, loose relations and problems with a deductive-nomological model

The wide representationist has two obvious replies to the above argument as follows: either there are functional correlations, and hence nomological laws, relating the variables associated with terms that constitutes the relation that determines the wide content of the mental state and a third event; or perhaps the deductive nomological law requirement is too strict a constraint on explanation.

Given a reliance on strict relational properties by the wide representationism, the first option of arguing for a nomological law between mental states and behaviour does not look plausible. But before the wide representationist rejects the deductive nomological model outright they might attempt to salvage the deductive nomological explanatory model by adopting a looser conception of the relation between the variables that are associated with the terms that constitute the law. The argument in the previous section is not so much that all relations can never be explanantia. The argument is instead that the relations that are appealed to by the wide representationist are not of a type that will provide for functional correlations between the terms of the putative law. But even if the variables associated with terms in a relation can change, is the relation itself

explanatory in the sense of furthering our understanding, or are functional correlations that might include relational variables merely descriptions of universal regularities? Must an explanation that increases our understanding go beyond the *relational* terms that constitute variables in a functional corrleation and the functional correlation in order to offer an adequate or deeper explanation?

The wide representationist might adopt a more liberal conception of relations, whereby variables associated with all the terms in a relation can change, and there still exists a legitimate functional correlation between the terms that constitute the law. But if any type of relation can never be a term in a functional correlation that offers an explanation in the deeper sense of understanding, then the deductive nomological account might be rejected on the grounds that it fails to offer a comprehensive account of scientific explanation perhaps if the laws are to be understood as equations that merely describe universal regularities.

For example, in a situation whereby the temperature of the first object gets hotter and the second object gets cooler, and the relation between the second and the first object is of 'becoming cooler' relative to the first object. There clearly could be, although this is by no means necessary, a functional correlation whereby the second object getting hotter stands in some functional correlation to the second getting cooler, one has only to think of a fridge to imagine the possibilities of certain correlations between temperaturechanges. As the fridge motor works harder, it gets hotter by a certain amount; this pumps more coolant around the fridge, thus cooling the air in the fridge a little more.

Another example that apparently demonstrates the explanatory role of relations is to consider a seesaw and the relative heights of the two people sitting on the seesaw. Arguably Fred's height relative to Jane explains the height of Jane. If Fred is lower than Jane, Jane is higher than Fred. If Jane is higher than Fred, Fred is lower than Jane. Fred being lower than Jane explains why Jane is higher than Fred. This appears tautologous, but there is a sense in which their relative heights are dependant on one another in a non-empty manner. Fred's specific height *relative to the pivot point*, which

through a rigid plank that sits on the pivot, explains Jane's specific height *relative to the pivot point*. It so happens that Fred's height *relative to Jane* is also a function of Fred and Jane's height relative to the pivot point. It does not really look like an appeal to the relation between Fred and Jane is that useful explanatorily, although this relation does serve to identify a variable associated with Fred and Jane that is statistically relevant - distance.

The question, however, remains as to whether the *relation* between terms explains functional correlations or distinct events or processes, or whether a change in the variables associated with the terms of the relation and underlying causal mechanisms deeply explain the functional correlations or distinct processes or events. It is not clear as to how the relation 'getting cooler' itself has any deep explanatory role in explaining *how* the fridge motor getting hotter results in air in the fridge getting colder, thus cooling the beer. A lonely appeal to the relation in the seesaw example looks downright vacuous. The functional correlations between the variables might describe a regularity, but perhaps the relation alone does not offer an understanding of the event or process to be explained.

But surely a relational property of distance constitutes one of the terms of the Newtonian Universal Law of Gravity; this relation is a crucial element of a fundamental or basic explanatory law; thus a relation can constitute a legitimate explanantia.

$$F=Gm_1m_2/r^2$$

(F is the force between two masses, G is the gravitational constant, m_1 is the first mass, m_2 is the second mass, r is the distance between the two masses.) In this case the distance between two objects is a key element in explaining the force between two objects. An altered distance between two objects will result in a different force between two objects. However, this example is arguably somewhat misleading as an attempt to justify that relational terms in a functional correlation can offer an explanantia with regards to an understanding of the nature of the 'force' between two objects of mass.

As Salmon (1999) plausibly argues, when discussing whether gravity provides an example of action-at-a-distance, there are two options for giving a deeper account of the nature of gravity as follows: either that what really provides an explanation for the force between two masses is yet to be discovered sub-atomic particles described as 'gravitons'; or on Einsteinian grounds a different conception of space as non-planar leaves the phenomenon of gravity as an unsurprising implication of an altered geometry. The property of distance is not the simple relational property that offers an explanantia for the 'force' between two objects with mass that Newton's law suggests. On either of the above attempts to provide a deeper understanding of gravity the relational property of distance does not provide an adequate explanation of gravitational forces in the sense of giving us an understanding of what explains the 'force' between two objects, despite providing what appears to be relational variable in a law that offers a description of how objects with mass behave.

According to the wide representationist, relational properties entirely determine the phenomenal character of mental states. It is simply not clear that the strict relations to which the wide representationist might appeal could fit into functional correlations with the events that the wide representationist hopes to explain. If the relational properties to which the wide representationist appeals, resemble the relation between the orientation of coins and particles, an example of a strict relation, then it is simply not possible that variables associated with the first term of the relation bear any functional correlation, and hence explanatory relevance, to variables associated with the second term of the relation or to a third distinct event. If there is no functional correlation between the variables associated with one term of the relation that constitutes the content of the mental state of an individual and the variables associated with intrinsic states of an individual, and thus no laws of nature relating these variables to the intrinsic states of the individual, and crucially the ensuing behaviour of the individual, then an explanatory model, following the Hempelian deductive nomological schema, cannot be constructed.

The wide representationist might adopt an account of intentional content that is constituted not by strict relations but by loose relations. But if there are functional correlations between the terms of the looser relation, the *relation* alone does not explain the changes in the variables associated with these terms, nor does the *relation* explain, in a sense that provides a deeper understanding, changes in any independent third event or process.[45] Thereforeon either a strict or a loose conception of relations, assuming only a deductive nomological model of explanation, wide intentional contents plausibly have no deep explanatory value concerning mental states and the ensuing behaviour of an individual.

But what about the prevalenceand success of generalisations that one might construct to explain the actions of agents in terms of beliefs and desires – sometimes known as 'Folk Psychology'? Perhaps the mere existence of these generalisations lends credence to the existence of nomological laws between wide contents and behaviour. These generalisations between beliefs, desires and actions might be constructed, but it is not clear either how such generalisations are derived from laws that identify specific functional relations between the relevant variables, or how they are deeply explanatory. If these generalisations can be constructed, and they rely on terms determined by strict or loose relational properties, then these generalisations are certainly not of a type that resemble generalisations derived from specific functional correlations between variables in more traditional laws of nature. It is perhaps illustrative to consider an attempted generalisation that relates beliefs, desires and actions.

If a believes that (b is the only (best, easiest or cheapest) means to achieving c) and a only (mostly) desires (c), then *ceteris paribus* a will do b.

By adding some specific antecedent conditions, a deductive nomological explanation might be provided. Bob believes that crossing the room to go to the fridge is the only (best, easiest or cheapest) means to getting a beer, and Bob only (mostly) desires a beer. The general law combined with the

---

[45] The next chapter considers specific representationist attempts to demonstrate how relations might be explanatory.

antecedent conditions; therefore, explains why Bob crossed the room to go to the fridge.

However, to restate the problem with the above appeal to folk-psychological laws, if these beliefs and desires are entirely constituted by strict relations, then beliefs and desires are not relations of a type that can plausibly constitute terms in a law because of the inability to construct a functional correlation between the variables (compare with Fodor's discussion of 'H' and 'T' particles). If beliefs and desires are only determined by strict relational properties, the functional correlations from which to derive these generalisations cannot be constructed, thus casting into doubt the explanatory power of the generalization.

If beliefs and desires are determined by looser relations, then this is consistent with the possibility of changes in the intrinsic properties of mental states and desires as a result of changes in variables 'at a distance' from those mental states. But then the question remains as to whether it is the relational property alone that is of deep explanatory relevance to the ensuing behaviour.

If, on the other hand, the beliefs and desires are determined not only by relational properties, but non-relational properties, then they might be associated with variables that fit specific functional correlations with ensuing behaviours. The above generalisations can be saved, but only by rejecting the claim that beliefs and desires deeply explain in virtue of only their relational properties.

The mere construction of generalisations relating beliefs, desires and actions does little to establish that wide contents and the determining variables that fix these wide contents can enter into law-like relations with neurological states and ensuing behaviour. If the deductive nomological model of explanation is adopted, then the difficulty in relating variables associated with relational properties, which supposedly determine wide contents, with the variables associated with events that the wide representationist hopes will be explained suggests a reason to be suspicious that wide contents can be explanatory in the way hoped for by the wide representationist.

However, what if the deductive nomological account of explanation is rejected?

Is the second option for the wide representationist, relying on an explanatory scheme that does not depend on laws of nature or functional correlations, plausible?[46] The wide representationist is not entirely without the means to reject a nomological law account of explanation - perhaps by appealing to various special sciences that make no appeal to strict laws or functional correlations between variables.

> [T]he question of whether there are folk psychological *laws* reflects in my view a misplaced emphasis on the roles of laws in scientific explanation. . . Paradigmatic examples of such laws are drawn from physics and chemistry. . . But if being covered by such laws is the mark of a scientific explanation, what of the special sciences? . . . [E]xplanations in the variety of special sciences, including *natural* sciences, such as geology and evolutionary biology, are perfectly all right whether or not there are laws that cover them (Wilson, 1997, pp.170-1).

This mere appeal to the legitimacy of the special sciences and specific examples within the natural sciences other than physics and chemistry, however, hardly offers an alternative set of explanatory standards by which explanation might be deemed legitimate or otherwise. Do these special sciences, or specific natural sciences, place *any* constraints on explanation? Is there a causal or explanatory 'free for all' within these sciences? Arguably any appeal to an explanatory schema entails some methodological commitments that provide some kind of constraints on explanatory relevance. Once the deductive nomological model of explanation is rejected there remains the somewhat pressing problem as to what replaces this model in different explanatory contexts. Whatever replaces the deductive nomological model might also present a problem to the wide

---

[46]Davidson, in attempting to maintain a consistent physicalist identity theory, takes the rather radical step of ditching explanation in favour of causal efficacy. Events, under a mental description, cause. Such events can also be subsumed under a physically description, but physical descriptions of these events are not of a type that can form terms in strict deterministic laws. However, it is not clear that content theorists want to lose explanatory relevance for content states by adopting this Davidsonian approach, although this remains a somewhat intriguing option.

representationist. Perhaps the special sciences, and incidentally some of the natural sciences, appeal to statistical explanation.

An alternative account of explanation is to appeal to a *high* probabilistic account. One event explains another if there is a high probability that it causes another event. If the probability is formulated as a probabilistic law, then the probabilistic explanation corresponds roughly with Hempel's (1965) Inductive Statistical model of explanation.

The earlier general psychological law might be stated in probabilistic terms:

If a believes that (b is the only (best, easiest or cheapest) means to achieving c) and a only (mostly) desires (c), then, a *probably* will do b.

By adding some specific antecedent conditions, an inductive statistical explanation might be provided. Bob believes that crossing the room to go to the fridge is the only (best, easiest or cheapest) means to getting a beer, and Bob only (mostly) desires a beer. Therefore Bob *probably* crossed the room to go to the fridge, and if he did, the high probability of the event occurring, explains why he crossed the room.

However, it is not clear as to how the variables associated with one of terms in the relation that determines the wide contents which putatively determines beliefs and desires has any *statistical relevance* to intrinsic states and ensuing behaviour, or whether strict *relations* have any statistical relevance; therefore, relations can not play a role even in a probabilistic law. Before this claim is developed further, it is worth considering how Salmon offers some substantial improvements on Hempel's inductive statistical model of explanation based on statistical relevance.

4.3 Statistical relevance and content

Salmon (1999) originally proposed a substantial revision of Hempel's Inductive-Statistical (I-S) model of explanation, by replacing it with a statistical relevance account of explanation. Statistical relevance is best explained in terms of an example. When considering lung cancer, one only

has to look for variables that have some (even small) statistical relevance to getting lung cancerin order to explain why an individual gets lung cancer. Smoking, for example, clearly has some statistical impact with regards to getting lung cancer. Notice that irrelevant variables that nonetheless have some statistical relevance have to be screened off such as the statistical relevance of yellow fingers (stained by the smoke) with regards to the increasedlikelihood of getting lung cancer.

According to Salmon, the structure of the universe is inherently statistical; therefore, explanations based on statistical relevance might be fundamental. He also argues that Hempel's I-S model is really only an epistemically impoverished version of his deductive nomological (D-N) model. Salmon also rejects Hempelian accounts of explanation on more specific grounds than a general appeal to special sciences, or specific natural sciences. Salmon argues that the argumentative structure that inspires the above non-causal, Hempelian deductive-nomological and inductive-statistical models, has no temporal component. Causality requires a temporal component; therefore, both the Hempelian D-N and I-S models are flawed.

Salmon has another convincing rejection of Hempel's I-S model. He rejects Hempel's I-S model because it violates explanatory symmetry: if high probabilities explain, then so must low probabilities. Scientific enquiry is often stimulated by the identification of variables that have some probabilistic, not necessarily high, but nonetheless statistical relevance to other variables. For example, the incidence of Leukaemia was higher amongst a population exposed to Nagasaki and Hiroshima. There need not be a high probability for these people to contract the disease, especially if they were a considerable distance away from the explosion, but there was a slight but increased probability, relative to the general population, that these people would contract Leukaemia. If high probability is the only criteria for justifying an event to be a cause, then the explosion, even when the probability is low, is not a cause of the Leukaemia. However, the explosion, even when the probability is low, clearly is causally relevant; therefore, high probability is not the only criteria for causal relevance (Salmon, 1999, p.130). All this suggests a constraint on an event being a cause that cannot be based on the level of probability, but whether the

occurrence of that event has any statistical relevance whatsoever to the occurrence of the other event. Therefore, this statistical relevance account also suggests a necessary constraint on explanation.

If one event, or variables associated with an event, has no statistical relevance regarding the occurrence of, or variables associated with, another event, then the first event is not a cause of the second event.[47] To return to the billiard ball example, it is not so much the fact that there are no functional correlations between variables, such as say the colour of the table and the motion of the billiard balls, which rules the colour of the table out as explanatorily relevant, but because the colour of the table is a variable that has no statistical relevance whatsoever to variables associated with the motion of the billiard balls. To return to Fodor's example, it is not because there is no law of nature between the orientation of coins and variables associated with particles in distant star system, which rules out the orientation of the dime being causally relevant. It is because the orientation of the dime has absolutely no statistical relevance to variables associated with particles in distant solar systems, that rules out the orientation of the dime as explanatorily relevant.

It appears that just as it is not necessary for the distinct terms in a relation to stand in any sort of functional correlation with one another, there is no need for distinct terms in a relation to have variables associated with them to have any sort of statistical relevance associated with the other term or a third distinct event. An event might occur a certain distance from Cambridge, but all sorts of variables associated with Cambridge need have no statistical relevance to variables associated with this event.

Not only do the terms in a relation need have no statistical relevance to one another, it is not clear how *relations* provide an explanatary understanding of ensuing events even when they appear to have statistical relevance to ensuing events.

---

[47] It might be possible to construct complex causal chains whereby statistical relevances cancel out, thus making it only appear that altered variables associated with one event have no statistical relevances to changes associated with variables associated with the other event. I will assume that such problems can be overcome.

To return to the relation 'taller', the relation itself has no statistical relevance to events other than through the variables associated with the terms themselves. Now it may so happen that variables associated with the two terms in a relation may have some sort of statistical relevance to each other (cf. the fridge), but this is not because of the relational property. For example, Bob might get the cookie more quickly that results in him growing a little bit taller, while Joe does not get the cookie, so he starves and shrinks. In this case it *appears* that Bob being taller enabled him to get to the food, which made him taller still, while Joe being shorter meant that he did not get the food and he shrunk slightly. However, an explanation that provides us with an understanding rather than a description of a regularity would be that it was the height of Bob that enabled him to get a cookie, which caused a change in Bob. The height of Joe explained why he did not get the cookie, and his starving provides us with an understanding of why he shrunk slightly.

But does the relational property of distance between two masses not have statistical relevance with regards to the force between those two objects? Distance has statistical relevance with regards to the motions of two masses. However, more basic laws involving yet to be discovered particles that do not rely alone on the relational property of distance might offer a deeper understanding of the gravitational 'force'. Alternatively, perhaps if the geometry of space is non-planar, distance loses its statistical relevance with regards to the 'natural' motions of objects. Gravitational force is 'geometrized' out of the picture; thus rendering this proposed counterexample to the lack of statistical relevance of relational properties as somewhat moot.

On Lombard's narrow definition of relational properties, it is in fact necessary that variables associated with one term in the relation undergo no change, other than with respects to the relational property, whilst the variables associated with the other term change; therefore, there is hardly room for statistical relevance between the variables associated with the terms. On a looser conception of relations, if the variables associated with the terms that constitute the relation do happen to have some kind of

statistical relevance, it is arguably not going to be the relational property that deeply explains this statistical relevance. Therefore, wide contents arguably have no causal or explanatory relevance in the deeper sense of being able to provide an understanding of the neurological states of individuals and these individuals' ensuing behaviour.

The wide representationist might reply either that some relational properties have statistical relevance, or that just as the D-N model was inadequate, the statistical relevance constraint on explanation is mistaken. It is not clear, however, how the strict relational properties that determine content can be of statistical relevance to mental states and ensuing behaviour. I suggest that the onus is on the wide representationist to demonstrate exactly how the variables associated with terms in a looser relation have statistical relevance to one another that offers a deeper understanding of the events or processes to be explained. This raises some perplexing problems for the wide representationist in how they argue that relations do have statistical relevance. Twin Earth scenarios certainly suggest that the statistical impact of the variables that constitute one term in the relation that determine wide content on variables associated with mental states and ensuing behaviour, is non-existent, if the variables associated with the states of the agent are by definition, *non-changing,* assuming the strict Lombardian conception of relations.

It is often assumed that the Twin Earth scenario is about the supervenience of mental states on intrinsic properties, and this explains why mental states and behaviours do not change. This is a mistaken conception of what the Twin Earth scenario is about; it is about *relational properties.* It is about how, on one conception of relations, variables associated with one term in a relation, other than the relational property, necessarily undergo *no change.* On this strict conception of relations, the relational change occurs only because of a change in the variables associated with the other term. The Twin Earth scenario is a somewhat stark illustration of this feature of strict relations.

If there are non-relational changes in the mental states of the behaviour of the earthling, then these changes are arguably what are going to do the

causing of any third event, and not the *relation*. If the variables associated with the terms that constitute the wide content relation change in a non-relational manner, and there is a mechanism that relates the variables associated with terms of the relation, then arguably this mechanism arises in a way that the wide representationist does not desire (more on this later). On both a strict and a loose conception of relations, assuming a statistical relevanceaccount of explanation it is not clear how wide contents are deeply explanatorily relevant.

There are, however, alternative causal explanatory models to which the wide representationist might appeal that reject the need for law-like functional correlations or even statistical relevance. These causal accounts appeal to counterfactual relevance.

4.4 Counterfactualism and content

To return to the billiard balls: there are various explananda that might be constructed given the billiard ball scenario, perhaps the most obvious being the event of billiard ball 'b' entering the pocket. This immediately suggests a certain categorization of explananda into the occurrence or non-occurrence of an event. The main focus of one type of explanation is the event of the ball falling into the pocket, in contrast with the event of the ball not falling into the pocket. This perspective leads quite naturally to what might be described as a counterfactual theory of causation. A cause might be proposed as the occurrence of one event that results in another event occurring as opposed to not occurring.

Imagine a slightly altered scenario and consider which changed events have absolutely no impact on the subsequent occurrence of the event of the ball falling into the pocket. These are the events, in the actual scenario, which we can claim are not causal with respects to the ball falling into the pocket. Furthermore, imagine a slightly altered scenario whereby one event not occurring, or being changed, results in the non-occurrence of the ball falling into the pocket. We might plausibly consider the occurrence of that event, in the actual scenario, to be a cause of the ball falling into the pocket. The table being a different colour does not look like a cause of the ball entering

the pocket. Why? Because, if we were to imagine an altered scenario where the table is a different colour, the ball still falls into the pocket. Ball 'a', moving in different direction, thus entirely missing ball 'b' looks to be important. Why? Because if we were to imagine the altered scenario, whereby ball 'a' misses ball 'b', then ball 'b' fails to fall into the pocket. It is worth pointing out the importance of keeping all other factors constant, whilst one considers the change in the event, we suspect to be causal. As a result of such considerations, one might construct a counterfactual principle of causation based on event occurrence. If there is a close possible world whereby one changes the occurrenceof one event, and the non-occurrenceof that event has no subsequent impact on the occurrence,or non-occurrence,of another event, then, that first event has no causal efficacy concerning the other event.[48]

If this general counterfactual principle of causation can meet all objections, and the wide representationist account entails that the variables associated with the terms that constitute the wide content relation, violate this constraint with regards to each other or certain subsequent events, then it can be inferredthat wide contents are not causally efficacious with regards to the variables that constitute the relation or subsequent events. Assuming the counterfactual account, or a good enough modification, does the wide representationistposition entail that contents are not causal? Consider once again the Twin Earth scenario.

On Twin Earth the Twinling has a mental state with twater content. On Earth, the Earthling has mental states with water content. If we imagine the Earthling on Twin Earth, the Earthling has a mental state with a distinct content. His belief is now about twater, yet the change, i.e., the change in

---

[48]It would be foolhardy to suggest that this general principle is foolproof without further need of modification. One counter example to this principle, sometimes presented, is the example of the firing squad. In the firing squad example there are six shooters, all contributing to the death of the victim. However, one can consider a close possible world whereby one removes one shooter, yet the victim still dies. Indeed one can consider a possible world whereby one removes each of the shooters, yet it does not follow that the individual shooters did not cause the death of the victim. I do not intend to dwell on the possible modifications that might be made to salvage this example, but the important point here is that there is a general principle of causation that puts constraints on what might legitimately count as a cause.

the chemical composition of water (and whatever other factors are deemed relevant), present to the Earthling, and hence the change in wide content of the Earthling's mental state, has no direct impact on the subsequent events that follow, regarding the Earthling's *mental* states and his ensuing behaviour. By Lombard's strict *definition* of a relational property, there necessarily is *no* change in the variables associated with events that form one term of the relation, in this case variables associated with states of the Earthling other than the relational property. There is a close possible world whereby the change in the chemical composition of water (and whatever other factors are deemed relevant), and hence the content of the Earthling's mental states, has no subsequent direct impact on the occurrence or non-occurrenceof the ensuing mental states and behaviour of the Earthling. A mere application of the counterfactual principle of causation entails, at least in this scenario, that wide contents are not explanatorily relevant to the Earthlings neurological states and his ensuing behaviour.

The representationist has various possible replies to this scenario as follows: first, to deny that the content of the Earthling has changed; therefore, the representationist is not committed to a change in content; second, to deny that the change in content has no impact on the subsequent behaviour of the Earthling; or third, to reject the counterfactual constraint on causation. I take the most plausible representationist variant of the first line to be an appeal to narrow content, perhaps as a function that maps contexts onto contents. However, I take it to be a distinctive feature of the wide representationist account that it is their consistent and strong affirmation that the wide content of the Earthling changes, and that it is only the wide content that counts. This essentially blocks this line of reply for the representationist. Besides, narrow contents have been rejected as a plausible account of phenomenal characterin chapter two.

Does the behaviour of the Earthling change? Perhaps the change in molecular structure of the water and hence a change in one variable that is one term in a relation that determines the wide content of the Earthling's mental states is responsible for a change in some variable associated with some state of the Earthling, other than the relational change, that results in distinct behaviour. The 'independence' of the two terms of the relation is

not analogous to a change in, for example, the colour of the billiard ball table and the event of the ball falling into the pocket. It is not, however, clear as to why changing contents is dis-analagous to the example of the change in colour of the billiard table and the ball falling into the pocket. It is precisely at this point that the burden of proof might be placed on the wide representationist to demonstrate how the change in molecular structure of water (and whatever other appropriate variables) does have some kind of non-relational impact on the subsequent behaviour of the Earthling.

One explicit wide representationist reply is that the Earthling reaches out for a glass of twater rather than water; thus his intentionally described behaviour is different on Twin Earth. However, these examples fail Fodor's (1994) cross context test. Additionally, a more substantial difference is required for the change in content to be really described as causal with respect to the *relevant* effects. Consider the billiard ball table again: by changing the colour of the table, there is a change in the effect, with regards to the event of the ball falling into the pocket: the ball rolls into a pink pocket instead of a green pocket. Yet clearly we do not want to claim that the change in colour of the table is a change in an event that is worthy of explanatory consideration *concerning the effect of the ball rolling into the pocket*.

Another possible consideration that might motivate a (non-relational) change in the Earthling's behaviour is to imagine the presence of a type of predator on Twin Earth and Earth. When these predators spot new arrivals to the planet that have drunk twater, they find these new arrivals particularly desirable to eat. These predators also live on Earth, but they never happen to chase Earthlings because they never find them tasty since there is no twater on Earth. After the Earthling takes a swig of twater on Twin Earth, that Earthling's neurological states and behaviour changes because of the chasing predator. However, the putative change in one term in the relation that determines the wide contents of the earthling's mental states is really of no *direct* explanatory relevance to the change in the mental states and the ensuing Earthling's behaviour. The change in the Earthling's behaviour only comes about incidentally because of the existence of the predators.

However, presumably the wide representationist wants wide contents to be explanatory in a way that the relations of mental states explain behaviour.[49]

The wide representationist might attempt to modify a counterfactual account that somehow rules out the Twin Earth example to be of any relevance to the real world. This is done by appealing to counterfactuals that have relevance only in nearby worlds and Twin Earth is a world a long, long way away.

> Generality is an explanatory virtue only insofar as it indicates counterfactual rigour, where we can express this in terms of an explanation's holding in *nearby* possible worlds. For the preceding reasoning to be valid, the set of Twin Earth scenarios just mentioned must form a subset of nearby possible worlds. Yet they do not: Many of the worlds in which doppelgangers exist . . . are *very* different physically from the actual world. . . [W]e could choose to compare Rex only to a physically very different individual who instantiates the same folk psychological states but exists on a planet that is otherwise *identical* to earth. In such a case we might be tempted to consider that the wide explanation is causally deeper. . . An exclusive or primary focus on doppelgangers can lead one to forget that in the actual world in which our psychological explanations are developed, the subjects of psychology are individuals who are not physically identical (Wilson, 1997, pp.211-2).

This is a popular line of reasoning. One might also point out that Twin Earth is a possible world that is more than far, far away; it does not even exist.

> A fact that tends to get lost in the excitement about our Twins is that *we have no twins*. Neither here nor on Earth nor anywhere in darkest space can molecule for molecule duplicates of flesh and blood human beings be

---

[49]Fodor considers a machine that tracks your past. This machine responds differently to you, depending on how your past went, so your past states combined with your present state are a causal power that can result in a difference in causal powers between two intrinsically identical beings; thus individualism is false. His response to this argument goes: "Individualism does *not* say, however, that having water thoughts rather than twater thoughts is not a causal power. What it says is that having water thoughts is not a causal power *in virtue of its being responsible for you producing water behaviours rather than twater behaviours*." (Fodor, 1995, p.213). Curiously Fodor tries a different tack in responding to the camouflaged moth-like examples (see his appendix in 1995). All he needs to do is consider the predators, or in his example 'the creatures you can sneak up on', as analogous to past history detectors.

> found. As an immediate consequence, the foregone generalisations [anyone intrinsically just like say you is in his circumstances going to do just as he does] are generalisations over things *all but one fail to exist.* This may not make the generalisations any less true, but neither does it recommend them as crashingly important (Yablo, 1997, p.260).

Yablo's attempt to diffuse Twin-Earth relies on the point that in the actual world people are physically very different and only wide contents provide the best explanation for explaining common behaviours rather than an attempt to identify shared intrinsic properties. Most importantly, claiming that identical people behave the same way offers no explanation for the behaviour. Yablo's specific argument will be considered in more depth in the next chapter. But for now, I am inclined to think that both Yablo and Wilson are missing the point of Twin Earth. Yes, we should not be looking for, nor should we necessarily care whether there are, molecular duplicates on earth, but there will be plenty of variables associated with psychological states that are determined only by a variation in a variable that occurs a long way away from the variables associated with that neurological state of an individual. Twin Earth merely serves to remind the wide representationist, somewhat vividly, that their theory relies on the explanatory relevance of *relations*. There is I think a clear sense in which thought experiments like Twin Earth do not really serve to establish the desired conclusion; i.e., Twin Earth does not demonstrate that variables associated with terms that constitute loose relations are *never* causally related, but Twin Earth does force a choice on the wide representationist who chooses not to avoid a decision as to which kind of relational properties they wish to adopt. Are they adopting a relation along the lines of Lombard's conception, or are they adopting a looser conception of relations that allows for non-relational changes in variables associated with the terms that constitute the relation? If strict relations are chosen, then the explanatory constraints are sharp. If the looser conception is chosen, it still has to be established how the loose relation does not violate any of the explanatory constraints without appealing to the variables associated with the terms that constitutes the loose relation or underlying causal mechanisms.

The wide representationist might entirely reject a counterfactual account and appeal to an account of causation that relies neither on nomological law-like relations, nor statistical relevance relations, but interactive causal processes.

4.5 Interactionism and content

Although ultimately disagreeing with the causal objection to representationism, Colin McGinn (1989) presents a conception of causation thus: "What happens at the causal nexus is local, proximate and intrinsic; the features of the cause that lead to the effect must be right there where the causal interaction takes place" (p.133). This causal limitation relies on what might be described as a principle of the locality of causation. Salmon's (1999, pp.248-60) latest proposed account of causation relies on causal interactions in a manner that also significantly emphasises the locality of causal interactions. The details of this account are problematic to the content theorist given the lack of mediatory processes between the terms that constitute the strict relations that determine wide content, whereby one of the terms is a neurological state. Conservation of momentum between two billiard balls best illustrates Salmon's principle. Salmon argues that the transmission of a 'mark' is required for any causal interaction. So, in order for one billiard ball to transmit energy to another billiard ball there must be some kind of local interaction, whereby energy is exchanged.[50]

The question for the wide representationist remains as to how relational properties transmit any energy. To return to Twin Earth, it is simply not clear how the altered chemical composition of water transmits any energy to the intrinsic states of the Earthling and the ensuing behaviour of the Earthling. If there is some casual interaction between the variables

---

[50] This involves a substantial defence of how certain effects at a quantum level do not really result in 'action at a distance' and various forces must be explained by exchanges of as yet undiscovered particles. I do not intend to embark on such a defence here, but it would be curious if the wide content theorist relied on a rejection of local interactions to support their position. Wilson (1997) explicitly employs the example of fields in an attempt to demonstrate how relational properties can be causal; i.e., distance relative to another object explains the behaviour of that object. Is he appealing to an endorsement of 'action at a distance'? If so, perhaps the onus is on him to explain how this is possible. It is certainly not clear that he has conclusive example of the causal efficacy of relational properties given the alternative options of altered spatio-temporal geometries or hidden particles.

associated with the terms that constitute the relation, i.e., some transmission of energy to the mental states of the Earthling, then there might be some grounds for maintaining that the altered chemical composition of water (and whatever other variables are deemed relevant) and hence the variables associated with one of the terms of the relation, can enter into a causal relationship with the intrinsic states of the Earthling and the resulting behaviour. This of course assumes that the relations are not narrowly defined in such a way that variables associated with one of the terms in the relation cannot change other than the relational property.

On a looser conception of relations, there could on Twin Earth be some causal mechanism via unusual predators (see earlier); however, these causal mechanisms are not of a type that is really relevant to the way in which the wide representationist arguably hopes the *relations* to be explanatory to an agent's actions. In the Twin Earth scenario there is no direct causal mechanism that mediates between the variables that determine the wide content and the mental states of the Earthling; therefore, the variables which determine wide content are not, in at least this one case, causally efficacious with regards to mental states and ensuing behaviour.

Burge, for one, has made considerable mileage from denying that any conclusion concerning how science should individuate its key terms follow from such metaphysical constraints, in specific local causation.[51] As stated earlier, whether Fodor's conclusion about the taxonomising of mental states follows from his explanatory principles is not a concern of this chapter. The concern of this chapter is whether an absence of functional correlations, statistical relevance, counterfactual relevance and local interactions between wide contents and mental events is sufficient to refute the wide representationist who claims that wide contents are explanatory.[52]

---

[51] "Local causation does not make more plausible local individuation" (Burge, 1986, p.16).

[52] There are complex issues concerning the individuation of mental states; how science ought to taxonomise terms; and the deep explanatory importance of such taxonomisation. Fodor argues that an appeal to intentional taxonomy in psychology results in laws that contain conceptually necessary truths, thus the laws cannot be causally interesting, such as 'if something is a water thought, it tends under certain optimal conditions to cause water related or water caused behaviour' (Fodor, 1995, op. cit.). Burge replies that the detail in such generalizations makes them both contingent and scientifically interesting. Burge provides an example

4.6 A summary of the constraints applied to two conceptions of relations

On the strict Lombardian conception of relations: Variables associated with the terms that constitute the relation that determines wide contents *necessarily* do not functionally correlate with variables associated with intrinsic mental states and ensuing behaviour. Variables associated with the terms that constitute the relation that determines wide contents *necessarily have* no statistical relevance with variables associated with intrinsic mental states and ensuing behaviour. Variables associated with the terms that constitute the relation that determines wide contents *necessarily have* no counterfactual relevance to variables associated with intrinsic mental sates and ensuing behaviour. Variables associated with the terms that constitute the relation that determines wide contents *necessarily have* no mediatory processes to allow for exchange of energy with intrinsic mental states and ensuing behaviour. Since the most plausible explanatory/causal models all require at least: a nomological law, statistical relevance, counterfactual relevance or intermediary processes, and the wide representationist fails to offer any other causal explanatory model that can account for the causal explanatory relevance of the terms that constitute the strict relation that determines wide contents with regards either to the terms or a third event hoped to be explained by the *relation*, strict relations do not, nor could they ever explain.

On the looser conception of relations: Variables associated with the terms that constitute the relation that determines wide contents *require* no functional correlation with variables associated with intrinsic mental states and ensuing behaviour. Variables associated with the terms that constitute the relation that determines wide contents *require* no statistical relevance

---

of the heart: 'if something is a heart, and it is functioning normally, it then pumps blood'. However, he then claims that such generalizations in physiology "[D]o not provide deep insight into causal relations. If physiology contented itself with such statements, it would certainly be remiss. But these [Fodor's] points do nothing to show that hearts are not taxonomically significant kinds" (Burge, 1995, p.233). This chapter is concerned with the role contents play in providing 'deep insight into causal relations' rather than the issue of taxonomic practise in science.

with variables associated with intrinsic mental states and ensuing behaviour. Variables associated with the terms that constitute the relation that determines wide contents *require* no counterfactual relevance to variables associated with intrinsic mental sates and ensuing behaviour. Variables associated with the terms that constitute the relation that determines wide contents *require* no mediatory processes to allow for exchange of energy with intrinsic mental states and ensuing behaviour. There may be all sorts of non-relational changes that occur in variables associated with the terms that constitute a relation which are consistent with functional correlations, statistical relevance, counterfactual relevance and mediatory processes. However, *relations* themselves do not change despite appearances to the contrary, and only appear to change because of the changes of the variables associated with the terms that constitute the relation. I would suggest that it is the changes in the variables associated with the terms that constitute *these* relations and underlying causal mechanisms that do the explaining in the sense that they further a deeper understanding of the events and processes. Furthermore, one can give an explanation only by explaining why the changes in the variables associated with the terms that constitute these relations do not violate the above explanatory constraints. Even on the loose conception of relations, relations alone cannot adequately explain in a manner that provides a deep understanding of processes or events.[53] Therefore, it can reasonably be concluded that wide contents have no deeper explanatory relevance to intrinsic mental states and ensuing human behaviour.

Hopefully this argument meets most of Wilson's methodological concerns. This argument does not rely on one, single principle generalised over scientific endeavour such as the existence of specific functional correlations between variables. It relies on a variety of constraints, including statistical or counterfactual relevance, which arguable equally applies to explanations within a variety of explanatory contexts. It does not rely on *a priori* constraints, just plausible pragmatic limitations that might be used to rule out certain variables from the outset of a scientific investigation. All of the

---

[53] Representationist attempts to explain how loose relations cause, will be considered in more depth in the next chapter

above might be described as the causal explanatory argument against the wide representationist.

One might construct a hierarchy of causal or explanatory principles starting with interactionism. This emphasises locality of causation. This might account for counterfactualism; why in counterfactual worlds there are events that have no impact on the occurrenceor non-occurrenceof other events. A further condition for having no causal relevance might be based on statistical relevance. An absence of interactions results in variables having no statistical relevance. Finally the identification of statistical relevances between variables might finally result in the identification of specific functional correlations between variables, which gives us nomological laws that describe universal regularities.

However, a preferred causal explanatory theory might abandon any of these constraints at various levels; for example, the requirement of nomological laws might be abandoned in favour of statistical relevance, especially if the events to be explained are fundamentally indeterministic, or if one is more concerned with application or policy, rather than underlying explanation. If one is trying only to explain the occurrence, or non-occurrenceof events in terms of the occurrence or non-occurrence of other individual events, counterfactualism might be adopted. If one is looking for low-level mechanisms, such as how traits are transferred between parents and offspring, interactionism might be the preferred model of explanation. All of this suggests a somewhat holistic approach to explanation where various models might be adopted depending on the explanatory context. What is curious, however, is that this holism is not sufficient for an explanatory 'free for all' with no constraints. Each approach will place constraints on what counts as legitimate explanation. I have identified these, above.

4.7 Conclusion

A dilemma is presented to the wide representationists: do they either adopt strict relations or loose relations as explanatorily relevant? If they adopt strict relations, they face serious problems explaining how the various terms that constitute that relation are explanatorily relevant with regards to each

other. These problems arise from four explanatory constraints considered above. The onus is on the wide representationists to explain why, if they want to appeal to strict relations, they need not meet the constraints of interactionism, counterfactualism, statistical relevance or functional correlations. A simple objection to any one of these explanatory models is not sufficient.

If they adopt a looser conception of relations, the wide representationist also faces problems. How does the wide representationist justify that their loose relations are explanatory in a way that provides a deep understanding of events and processes that strict relations do not? They will arguably have to appeal to the variables associated with the terms that constitute the relation and show how they do not violate the above explanatory constraints by demonstrating the undelying causal mechanisms. The looser conception of relations allows for the terms that constitute the relation to be explanatory in a manner that meets the explanatory constraints; however, it is not clear, even on the looser conception, how the relation itself contributes in any way to a deep explanation that provides understanding rather than mere descriptions of universal regularities.

However, perhaps full justice has not been done to the specific wide representationist attempts to join the explanatory club. The next chapter considers in more detail the wide representationist appeal to loose relations as an attempt to salvage wide causation; however, the wide representationists only rebut a straw anti-wide representationism. All their points entirely fail to address a stronger anti-representationist position. Straw anti-wide representationism maintains that no relational properties can be deeply explanatory. The real anti-wide representationist position maintains that there is a sense that relational variables could meet explanatory constraints, but the representational properties that consititute these relational variables can never be deeply explanatory.

Chapter 5: Wide Representationist Attempts to Join the Explanatory Club

5.1 Introduction

Either the representationist adopts an account of strict relations as having a causal or explanatory role or they adopt a loose conception of relations. If they adopt a strict conception they face serious problems confronting various plausible constraints on explanation. If they adopt a loose conception, they have to justify why these relations are explanatory without appealing to the variables associated with the terms that constitute the relation or underlying processes that provide an understanding of the relation. But given this dilemma, how can philosophers claim that wide contents are explanatorily relevant? One strategy appears to be to set up the anti-wide representationist position in a manner that is hardly conducive to plausibility. Once the absurdity of what might be described as straw-individualism is demonstrated, this putatively leaves ample room for wide contents as explanatory.

First, the anti-wide representationist might be characterized as holding that the variables associated with the terms that stand in certain relations to one another are *never* causally efficacious with regards to each other or a distinct third event. However, as discussed in the previous chapter, the anti-wide representationist is not committed to maintaining that variables associated with the terms of a relation are never causally efficacious if a looser conception of relations is adopted. The anti-wide representationist merely needs to maintain that the terms of the relations associated with wide contents are never directly explanatory with regards to intrinsic mental states and ensuing behaviours. If the variables associated with the terms do have, for instance, statistical relevance with regards to each other, then it is not the relation that explains in the sense of providing an understanding of the statistical relevance.

It is a feature, which will be developed further, that the kinds of examples usually given to demonstrate the causal efficacy of wide contents rely on both the looser conception of relations and 'third party perceptual

mechanisms'; these third party perceptual mechanisms bring together normally unrelated variables, making *apparently,* non-local, statistically irrelevant and counterfactually redundant variables associated with distinct terms in a relation both interactive, statistically relevant and counterfactually relevant. Therefore, these variables, associated with the terms that constitute the relations that determine wide contents, might indeed be deemed to be causally efficacious, even with respect to intrinsic mental states and behaviour. But the variables associated with the terms are not causally efficacious in a manner direct enough to satisfy the hoped for explanatory role for the variables associated with these wide contents on mental states and ensuing behaviour. By 'direct', I mean in the sense that the relation itself explains the statistical relevance between the terms that constitute that relation, or a third distinct event. Thus, the anti-wide representationist is not committed to claims such as 'the environment *never* matters'. Appeals to third party perceptual mechanisms can be posited which make terms that stand in a specific relation such as properties of an environment, as well as causal histories highly relevant. But, these indirect mechanisms provide explanatory relevance in a manner unsuitable for the role wide representationists want to place on relations as explanatory in the sense of providing an understanding of the events to be explained.

The anti-wide representationist might be accused of entailing the absence of the existence of indirect causal mechanisms. There may be complex causal processes that result in the variables that are associated with the terms, which constitute the appropriate relation, being explanatorily relevant. But once again, the anti-wide representationist need not deny the existence of complex causal mechanisms, but these mechanisms are not conducive for the explanatory role of wide contents hoped for by the wide representationists who maintain that phenomenal character is entirely exhausted by a wide content relation, and that phenomenal character is explanatorily relevant to both neurological states and the ensuing behaviours of organisms.

Finally the anti-wide representationist might be accused of entailing that the best explanation always appeals only to microphysical properties. But it is not clear why the anti-wide representationist needs to entail such an account.

The anti-wide representationist is only committed to a variety of methodological constraints on explanation in a variety of explanatory contexts. Arguably the best explanation involves a working out of how what appears to be a violation of methodological constraints on explanation does not violate methodological constraints on explanation. Through this resolution of how variables do not violate methodological constraints, more complete, deeper explanations are offered that do not appeal to even the looser conception of relations.

5.2 Relations that matter and the relations that do not

Some kinds of relational properties, for example, the relation of a particle in Alpha Centuri to the orientation of a dime on earth, are clearly relational properties that fail to meet the above explanatory constraints, and thus should be ruled out as explanatorily relevant; however, this, according to the wide representationist, ignores the relational properties that might matter. *These* relevant relational properties may be the most theoretically appropriate and the deepest explanatory properties. Here we have an immediate appeal to a looser conception of relations.

However, an immediate question arises as to how the relational properties that are supposed to be causally efficacious with regards to certain specific effects are distinct from the relational properties that clearly are not, with regards to these certain specific events. Any such distinction between these distinct types of relational properties appears to be grounded by the very same kinds of methodological considerations outlined above. Furthermore, any such attempt to determine why the variables associated with the terms that constitute the relation are causally efficacious, with regards to each other, by appealing to the above explanatory constraints, arguably results in the relation certainly not being the causally deepest explanation, or even necessary for the explanation.[54] If the wide representationist fails to

---

[54] This point might be justified on grounds similar to Kim's (1993) causal exclusion principle. It seems that the event to be explained by the relation is overdetermined; either the variables associated with the terms that constitutes the relation explain the event; or the relation explains the event; however, the event cannot be caused by both. Surely the variables associated with the terms that constitute the relation are better candidates for explanation.

demonstrate why their relational properties are distinct from the mere-Cambridge properties, with regards to certain specific events, then the wide representationist faces the challenge of meeting the above explanatory constraints.

Two types of examples will be presented on behalf of the wide representationists as an attempt to show how wide contents might be explanatory; both appeal to a looser conception of relations. The first type of example appeals to explanations in biology as an attempt to show how relational properties can be causally efficacious (Wilson, 1997, pp.192-3); however, I will argue that although there is a sense in which these relations might be deemed to be explanatory in a somewhat superficial manner, they only defeat a straw anti-representationism. The second type of example appeals to explanations in psychology (Yablo, 1997, pp.265); however, this example underestimates the relevance of shared intrinsic properties, so the example can not do the work for which the wide representationist hopes.

5.3.1 Moths and environments

One explicit attempt to construct an explanatory scheme that appeals to wide contents is to make reference to explanations in biology, specifically, explanations that involve interactions between creatures and their natural environments. Perhaps the kind of wide explanatory schema prevalent in these explanations is also applicable to wide contents (Wilson, 1997, p133). One example arises from consideration of the peppered moth.[55] The colour of the environment might be both statistically relevant to the survival of a moth, as well as counterfactually relevant to the survivability of a moth.

If the anti-wide representationist is committed to holding that environments *never* matter, then actual examples like this might present an immediate challenge to the anti-representationist whereby the colour of an environment is both statistically and counterfactually relevant to the survivability of moths.

---

[55] I am borrowing this example from Walsh (1998, p.632).

Consider the example of the camouflaged moth. Place the moth in a changed environment and its survivability decreases. Place the moth in its natural habitat and it will have a far greater chance of survival; therefore, relational properties, such as the *sameness* of the moth's colour relative to the colour of its environment, is explanatorily relevant to survivability of moths. This kind of example might also be presented as a suggestion as to how a creature in a distinct environment, despite having identical intrinsic physical properties, will have different causal powers. Thus the above Twin Earth examples do little to establish the anti-wide representationist conclusion (cf. with the predator example in previous chapter). Perhaps wide explanations also have a role to play in psychological explanations of behaviour. But this example of the explanatory relevance of relations only defeats a straw anti-representationistposition.

The moth example in no way violates the explanatory constraints outlined above. This immediately raises a question as to how the moth example salvages the hopes of the wide representationists who claim that explanations in biology will somehow suggest a *new* explanatory schema that accounts for the explanatory role of wide contents. When considered closely, the moth example does little to offer a new methodological constraint on explanation that meets wide representationist expectations.

The moth example makes the 'sameness' relation between the colour of the environment relative to the colour of the moth and the survivability of moths *appear* explanatory only because there are distinct variables associated with the terms that constitute the relation of sameness; and furthermore, there exists predatory organisms that have certain discriminatory mechanisms that detect when certain objects stand out from the background. The underlying mechanism gives us an understanding of why the relation offers a description that is explanatory in a superficial sense. One might argue that the predator is sensitive only to differences in colouring and not to the individual colours of the moth and the environment separately. But a further question remains as to whether this type of detecting mechanism proves that relations, of a type that the wide representationist hopes to appeal to, can be explanatory in a manner that provides an understanding of the events to be explained.

When a predator eats a moth, this is hardly a 'wide' explanation for why the survivability of the moth changes. A predator eating a creature also has a somewhat obvious causal impact on the reproductive potential for that creature. Additionally, the kind of wide explanations cited, such as the colour of an environment relative to the colour of the moth, is not clearly explanatory in a way that the wide content theorist can apply to wide contents in psychological explanation because this example relies on a third party detecting mechanism that provides the instantiating causal mechanism.

Is there some causal mechanism that explains the change in survivability of the moth when its environment changes? There are local causal interactions, which explain the decreased survivability of the moth. Consider the properties of the predators' perceptual systems. Further mechanisms can be spelt out involving reflectance properties of the moth's wings and the bark of the trees, including the causal mechanisms leading to the predators behavioural response, when it spots, and eats, a moth. However, there are no *direct* causal mechanisms between the colour of the environment, the colour of the moth and the moth's survivability. The causal mechanisms run indirectly through the variables associated with the terms that constitute the relation and the perceptual systems of the predators. The change in the colour of the environment is clearly a variable that is statistically relevant to the survivability of moths. But, the change in the moth's survivability is only indirectly because of the change in moth's environment. The survivability of the moth in the new environment does not change because of the relation between the colour of the moth and its environment, but because of the change in variables associated with the colour of the moth, its environment and the underlying explanatory mechanism; i.e., a change in the behaviour of the predators. It is a further question as to the role relations play in the detecting mechanism of the perceptual systems of the predator.

Concerning the survivability of the moth, there are counterfactual implications of a change in colour in the environment for the moth, but these also arise only because of the existence of the predators. If the colour

of the environment is a variable that changes relative to the colour of the moth, then one event, namely the chances of survival for that moth, also changes. However, this is so partly because of the presence of the predators. To the extent that the survivability of the moth changes, it is not directly because of a change in a relation, but because predators can now detect the moth, only incidentally because of the change of colour in one term that constitutes the relation; i.e., a change in the colour of the environment.

However, it might be claimed that all we have here is a difference of opinion about what variables we consider to be background assumptions, and which variables are explanatorily important. If we keep the predators constant, but change *only* the colour of the environment relative to the colour of the creature, then the change in the environment is more important, and the sameness relation captures this change. After all it is more likely that a moth flies to a new environment, or an ecological disaster changes the environment, with the continued presence of the same predators, rather than predators suddenly achieving finer discriminability in their perceptual systems. When explaining the colour of the moth, it is surely the colour of the environment that is relevant, instead of the capacity for discrimination in the perceptual mechanisms of predators. But, to identify only one counterfactually relevent relational property, which ultimately relies on a change in intrinsic properties of one of the terms of the relation, ignores the other possible counterfactually relevant variables that might eventually lead to a deeper understanding. Single counterfactually relevent variables hardly provide a deeper understanding of the events to be explained, but instead a suggestion as to how descriptive laws might be constructed.

To claim that relations offer a deeper explanation in the sense of providing us with an understanding of events is surely a mistake. If it is claimed that the colour of the environment, which is one variable associated with one term in the sameness relation, is the deepest explanation for the survivability of moths, this ignores the possibility of changes in predatorial discriminatory systems. But then it might be claimed that placing the emphasis on the predators as offering the deepest explanation ignores the colour of the environment. Not so! Two terms are important to the

'sameness' relation, both the colour of the environment and the colour of the moth. Furthermore, the importance of both the colour of the moth and the colour of the environment is subsumed under an analysis of the discriminatory capability of the predators while a single emphasis on one variable associated with the environment entirely ignores the relevance of the predators. The more complete the explanation, in terms of statistically relevant variables identified, the deeper the explanation. A lonely explanatory appeal to a relational property, which indirectly relies on intrinsic changes in one of the terms of the relation, combined with a functional correlation, at worst, ignores methodological constraints on explanation and offers, at best, incomplete explanations.

It is a further aspect of the above example that there is some perceptual apparatus that brings together variables, as terms in a relation, which initially appear to bear no relevance to one another. To restate, if there are no predators, then the relation between the colour of the environment and the colour of the moth really has no bearing on the survivability of the moth. To return to Twin Earth, if there is some detecting mechanism within a creature that can track causal histories of Earthlings and finds Earthlings tasty when they eat twater, then variables that really have no relevance can be 'causally linked' and made relevant. Even in the example of the colour of the billiard table apparently having no relevance to the ball going into the pocket, one can tell a story whereby the first ball happens to be the same colour as the table; therefore, the, as yet unmentioned, player fails to clearly distinguish the first ball from the table, miss-hits the first ball, causing the second ball to miss the pocket.

All these attempted solutions, which demonstrate that variables associated with wide contents might have some kind of explanatory relevance, rely on a third party detecting mechanism. Surely the wide representationist hopes that variables associated with these wide contents have a far more direct explanatory role to play concerning mental states and ensuing behaviours. These examples do show that the anti-wide representationist should not attempt to argue that the variables associated with the terms that constitute a relation, more *loosely* defined, are *necessarily* causally and explanatorily redundant with regards to *every* other event. This would indeed be a

mistake; this would be a straw-individualism that the wide representationist often takes great glee in demonstrating to be false. But this is not the real anti-representationist position. The real anti-wide representationist need only argue that even the looser relations which the wide content theorist appeals to are explanatory *poor* for explaining the events for which the wide content theorist wants these wide contents to be explanatorily useful.

However, the above argument partly assumes that the predators are not part of the wide explanation. What if the predators are also part of the wide explanation?

5.3.2 Moths and chameleons

> Explanations citing the selection pressures causally responsible for a given behaviour are wide, since selection pressures - such as the relative abundance of competing species, the presence of particular environmental toxins, and the existing physical members of the species - do not supervene on the intrinsic physical properties of individuals (Wilson, 1997, p.195).

Robert Wilson suggests that the existence of 'competing species' might be part of a wide explanation, so he would, I assume, take the existence of predators to also be part of a wide explanation. But this is a straw analysis of 'wide'. The anti-wide representationist can account for the collision of billiard balls as part of the explanation in terms of some kind of energy transmission between billiard balls that relies on intrinsic properties, but in no way claims that properties associated with the other ball are 'wide'. It is, thus, not immediately obvious how a predator eating a moth is a 'wide' explanation.

However, to do full justice to Wilson's account, it is worth developing further his own example of a causally efficacious relational property. In the first biological example, I took the attempted appeal to the colour of the environment relative to the colour of a creature and the survivability of the creature, independent of the intrinsic properties of a creature, as part of a justification for the causal efficacy of relational properties. Wilson instead of focussing on specific variables associated with the environment, such as colour, and the relation of sameness of colour to a creature, suggests that

degrees of specialization is relative to other creatures. And *this* relational property is causally relevant.

> Highly specialized species tend to extinction in times of ecological catastrophe. Roughly, a species is highly specialized if it adopts a limited range of survival strategies relative to other competing species. . . Being highly specialized is a property whose possession is causally responsible for species-wide extinction in certain circumstances: *It* is a causally efficacious property (Wilson, 1997, p.124).

Instead of just the moth, also consider the survivability of chameleons, which change colour depending on their surrounding environment. Once again the colour of the environment changes due to ecological change. Wilson is then claiming that, for example, chameleons are *more* adaptive to their environments; therefore, survivability is only relative to the survivability of competing species. This is acceptable, but is it really a relation that is the deepest explanation that provides us with an understanding for why a species tends to extinction?

First, it must be assumed that the survivability of some of the individuals belonging to a more adaptive species is a variable that has some statistical relevance to the survivability of the individuals belonging to a less specialized species. Wilson explicitly acknowledges in the above quote that the individuals within the relative species are competing. Is this assumption a relational property? Is this an assumption that can be ignored when assessing Wilson's general claim as the deepest explanation? Certainly the assumption that the individuals within the two species are competing relies on interactive causal mechanisms. Assuming the survival of some creatureaffects the survival of another creature, a story can be told. In the changed environment, the hungry predator fails to spot the chameleon, but it spots the moth, and so it eats the moth. See the above explanation to explain why the predator spots the moth when the environment changes. The chameleon reproduces because the predator does not catch it and the 'baby boom' chameleons eat all of the moths' food, thus the moths starve to death. Or one of the profligate chameleons spots the moth, successfully sneaks up on the moth, and subsequently eats the moth. There appears to be nothing relational about these instantiating explanations.

The wide representationist is thus far attacking a straw anti-wide representationist who maintains that distinct terms in a relation are never causally efficacious with regards to one another, and environments never matter. It is simply not clear that the anti-wide representationist is committed to such a position. However, the anti-wide representationist *is* committed to claiming that the relational properties never explain changes in the variables associated with the terms that constitute the relation or distinct events in the deep explanatory manner hoped for by the wide representationist.

5.3.3 Historical functions

When considering the colour of the moths, should one explain this, in terms of the chemical composition of the melanin within the moth, or in terms of the colour of the moth relative to its environment, or in terms of the successfulness of the trait in previous ancestors? The previous example has already shown that the colour of the environment relative to the colour of the moth might be statistically relevant to the survivability of the moth. Given certain evolutionary principles, such as the inheritability of traits and 'fitness' of previous generations of moths, the colour of the moth might be explained in terms of the previous generations of the moth. There are thus three possible explanations for the colour of the moth. First in terms of its current function, i.e. its colour helps it to be camouflaged from predators: second, in terms of it's historical function, its colour is a result of a selected trait that has been successfully selected for in its past: Third, in terms of the chemical properties of the wings of the moth and the reflectance properties of the wings of the moth.

The best explanation, concerning the colour of the moth probably depends on the manner and the context in which the explanandum is constructed. If someone asks why objects such as, for example, the wings of this moth look grey, one might explain this in terms of chemicals and reflectance properties. If someone asks, assuming a very recent ecological change, why there are more grey moths than white moths in this area, then, the current survival advantage to being grey rather than white might be the appropriate

explanation. More grey moths successfully hid from the predators. If someone asks why there are more grey moths, given a constant environment over a substantial period of time, or a change in environment many moth generations previously, one might explain this either in terms of the current survival advantage of being grey, or how this was also applicable to fitness in the past and this trait was selected for through the evolution of previous generations.[56]

I'll take the previous discussion of survivability of moths within a current environment to be adequate in analysing why current functions do little to establish the explanatory relevance of the relations that determine wide contents, perhaps historical functions do a better job. Once again, however, historical functions do not violate the methodological commitments outlined at the beginning of this chapter. A relation between events can be constructed historically. These events may even have variables associated with them that have statistical relevance to variables associated with later events. But, it is in explaining the statistical relevance between the events or the variables associated with these events and the underlying causal mechanisms that the deeper explanation is offered.

5.4 Levels of explanation

Another strategy the wide representationist might adopt is to appeal to distinct levels of explanation. An appeal to a higher autonomous level of explanation might be employed to rule out more detailed explanations, for example, those developed above as too detailed, while the more general explanations appealing to only the relations might be more appropriate. Wilson attempts to make a distinction between mere-instantiating explanations and higher-level explanations. The kind of explanations offered above that develop the causal mechanisms are mere-instantiating explanation.

> Sometimes an instantiating explanation, while specifying details about the actual causal processes by which an event occurred or identifying the intrinsic properties of some

[56]See Walsh (1998) for a nice distinction between current and historical functions and how the best explanation might depend on the interests of the investigators.

> cause, is not essential to a broader understanding of *why* that event occurred. . . I shall refer to instantiating explanations that are inessential to a broader understanding of the *explanandum* as *merely instantiating* explanations. Sometimes lower-level explanations are merely instantiating explanations, and, when they are, corresponding higher-level explanations provide the causal information relevant for explanation. . . I am not implying that there are cases in which one offers a better causal explanation for a given explanation by *avoiding* talk of causal mechanisms altogether. Rather the claim is that the most informative level of description at which an explanation can be pitched need not be the one that provides the greatest amount of causal detail. In explanation sometimes less is more. After all if this were not the case, the best explanations that one could offer for *any* phenomenon involving material objects would be microphysical, and that is absurd (Wilson, 1997, p.128).

Wilson suggests a *reductio* against the anti-wide representationist, claiming that in order to offer the deepest explanation, the anti-wide representationist is committed to going smaller and smaller. Hilary Putnam (1975b, pp.295-7) illustrates the problem with an appeal to the micro-physical as the best explanatory strategy with his example of why square pegs do not fit into round holes; the moral being that there is no point going microphysical to explain why pegs of a certain shape do not fit into spaces of a different shape.

However, the anti-wide representationist need not accept that smaller is always better, merely, that if explanations are offered, which apparently violate the methodological constraints outlined in chapter four, the explanations completeness might be improved by demonstrating how the supposed explanation does not violate these explanatory constraints. Because of apparent violations of methodological constraints apparent in Wilson's examples, they are always going to be substantially flawed in terms of completeness. I would suggest that Wilson's 'deeper' generalizations that appeal only to loose relational properties miss the importance of the methodological constraints outlined in the previous chapter. Only by meeting the explanatory constraints, will his relations be clearly distinguished from the strict relations he is so eager to reject. An appeal to the variables that constitute the terms of a relation is not a choice - it is a necessity, in order to provide an adequate explanation. The changes in relations cannot be divorced from the changes of the variables in the

terms that constitute the relation and the underlying mechanisms that provide a deeper understanding of the events to be explained.

To illustrate further, another example of a relational property might be considered that is analogous to Wilson's example of relative specialization. Consider two billiard balls heading, from different directions, towards a pocket that can only hold one ball. One might explain why one billiard ball eventually sits in the pocket, by appealing to relational property: 'the faster ball, gets the pocket'. Is this generalization really an appeal to a relational property? If it is, one might immediately ask whether it violates any methodological constraints. In terms of covering laws there is a simple equation that relates velocity with distance and time. It is immediately obvious that what initially looks like an appropriate deeper explanation ignores distance as an important variable. Besides, surely it is the properties associated with the objects that constitute the relation that explain why one ball gets to the pocket and the other does not.

At time t1 there is a specific distance that those two balls are from the pocket. If, at any time after t1, one ball is in the pocket, then the other ball cannot enter the pocket. Factoring in specific velocities and distances one can identify which ball gets to the pocket first. There are two balls with distinct causal paths that result in one of the balls sitting in the pocket at t2. These causal paths, however, intersect. The first ball to arrive at the pocket has some influence on the causal path of the other ball when the slower ball hits the pocket with the ball sitting in it. Without some interaction, the relative speeds of the balls have absolutely no relevance to the causal paths of the two balls. For example, the velocity of the rotation of my dime has no relevance to the rotation of one of Jupiter's moons despite the fact that one of the objects rotates faster than the other. Wilson needs to appeal to the variables associated with the terms that constitute his relations and the underlying causal mechanisms in order to motivate his distinction between strict and loose relations. Incidentally, it is the local interaction that makes the distinct properties associated with the balls relevant to the causal paths of the two balls, not the relational property.

On consideration, what initially appears like a legitimate appeal to a relation as explanatory far from offers a deep explanation. Describing one ball as the faster and the other as the slower is a convenient way of distinguishing between what otherwise appear to be two balls with identical intrinsic properties; it also serves to roughly quantify specific variables associated with the two balls. Assuming the distances from the pocket are the same when the two balls start, one ball having the greater velocity does have an effect on the other ball. However, a more complete, and dare I say better and deeper, explanation involves an identification of all the relevant variables applied to the objects that constitute the relation. Completeness does not always rely on the microphysical; however, it arguably does rely on laws, and/or statistical relevance, and/or counterfactual relevance, and/or interaction. A mere appeal to an individual *relation* will probably result in the failure of identifying all the relevant distinct variables associated with the terms that constitute that relation and at least in the above instance ignore the importance of some underlying causal mechanisms or local interaction.

If one individual within a species is more adaptive than another individual within a distinct species, then given a change in environment that individual belonging to the first species may have greater survivability than the individual within the other species. But the more adaptive individual is going to affect the individual within the other species only if the individual within the other species is competing for shared resources, or preying on that individual. Does a relational property explain this? No. Only instantiating mechanisms explain how individuals within one species might impact on individuals within another species and this non-relational assumption is essential to Wilson's general claim.

I would take issue with attempted claims that explanations in biology are wide or *only* appeal to relations, they certainly in no way violate methodological constraints on explanation, yet it is a feature of strict relations that they violate the above explanatory constraints. The appeal to looser relations ignores the importance of the variables associated with the terms that constitute the relation and the underlying mechanisms. An appeal to the efficacy of wide relations, however, need not only rely on

biological analogies. Stephen Yablo (1997) attempts to provide an account whereby explanations within psychology might best appeal to wide relations.

5.5 Recognising pictures

> When a modicum of extrinsic detail buys up an abundance of intrinsic, we have wide causation pure and simple (Yablo, 1997, p.372).

But how is this to be construed so as to be consistent with the somewhat plausible argumentation going on above which apparently demonstrates: first, that the *strict relations* which determine wide contents are neither causal nor explanatorily relevant because they necessarily violate explanatory constraints; second, if the variables associated with the terms that constitute the *looser relation* are of explanatory relevance to one another, it is not the *relation* which provides a deeper understanding of the relevance; if there is a third event, it is not the *relation* between the variables, associated with the terms that constitute the relation, which deeply explains the distinct, third event?

It might be argued that offering an explanation of behaviour in terms of over-specific intrinsic properties is a poor alternative to offering an explanation in terms of wide contents. This reply is consistent with the adoption of a loose conception of relations whereby different beliefs explain different behaviours on the ground of differences in the relations, but this is consistent with a change in the variables associated with the terms that constitute the relation. This, broadly speaking, is Yablo's suggestion. He offers an account of causation based on proportionality. According to Yablo (1997, pp.258-9), there are two ways to go wrong in explanation: the first is to over-generalise, e.g. matter, *ceteris paribus*, conducts electricity, and the second is to under-generalise, e.g., pennies, *ceteris paribus*, conduct electricity. To get it right is to identify just enough of the relevant variables to explain something, too little and the explanation is under-generalised, but if too many variables are identified, and you are left with a false over generalisation.

Yablo argues that adjusting intrinsic properties can affect behaviour while leaving intrinsic properties alone do not. However, he goes on to argue that this is *compatible* with the possibility of varying intrinsic properties a substantial amount whilst leaving behaviour intact. Additionally if the wide content of mental states is kept constant, then in the closest departures from reality behaviour will stay constant.

> If we distinguish "what I believe/desire" from "how I believe/desire it" as factors in extending my hand, then adjusting the how-factor alone *can* affect my behaviour while adjusting the what-factor alone *cannot*[57]. . . But this is fully compatible with saying that many or most ways of mucking with my [desire][58] leave my behaviour in place, provided that I keep wanting water. And it is supremely compatible with the notion that I would still have extended my hand if I had wanted water in the way involving the least possible departure from actuality (p.265).

To better illustrate Yablo's (1997) account of wide causation it is perhaps best to consider his own illustrative example. It is problems with his example that demonstrate the flaws in his reasoning. A child when looking at a picture of its mother responds with a happy gurgle. Given an identical picture, the child responds with a happy gurgle. If one makes the picture fuzzy enough, then the child will not respond. This, Yablo contends, is not what is interesting about children responding to pictures, yet this, according to Yablo is *all* that the anti-wide representationist has to say about the scenario.

> Adjusting the how-factor alone *can* affect Isaac's behaviour - had the photograph been much fuzzier Isaac would have been baffled by it - while adjusting the what-factor alone cannot - leave the colours alone, and regardless of the subject, Isaac grins (p.265).

---

[57] Notice that Yablo accepts the point that identical duplicates behave the same way, but he thinks that this is irrelevant.

[58] The brackets are his. He distinguishes between wide contents and non-extrinsic properties, respectively, in terms of *what* one believes/desires and *how* one believes/desires. He employs square brackets to distinguish a shared set or shared intrinsic property, or, in other words, an attitude with all the extrinsic aspects bracketed away; Yablo describe bracketed contents as battitudes.

Notice that Yablo only identifies colours as the relevant intrinsic properties. Far more explanatorily relevant, according to Yablo, is that if the child is given a *hugely* different variety of pictures, which are still depictions of its mother, the child still responds with a happy gurgle.

> Isaac is a boy capable of tracking his [mother][59] through a huge variety of photographic images, and the image at issue here is not anything special or strange but the one his [mother] *would* have given rise to if the actual image were for some reason ruled out (p.265).

Yablo then asks how many of the intrinsic properties of the picture have to be changed to extract a happy burbling response from the child. He contends that one could change the intrinsic properties a relatively *small* amount in order to make the child not recognise its mother, and yet one could change the intrinsic properties of the picture a *substantial* amount, and still elicit a happy response from the child. Very different pictures and poses of the child's mother bring out the same behavioural response:

> Why Isaac should lose sight of his [mother] in the alternative-image world nearest to this one is hard to understand. Harping on the fact that a change in intrinsic properties is necessary and, if suitably dramatic, sufficient for a change in Isaac's reaction only drives home the problem home: why should there be a *dramatic* change in the *nearest* alternative-image world to actuality (p.265)?

To restate Yablo's criteria for a good explanation, he maintains that just claiming that an identical picture will produce the same behaviour in the child is true, but does not really explain much. Changing the intrinsic properties of the photograph enough can effect a change in the behaviour of the child, but this is not *enough* to explain the behaviour being the same when a variety of pictures are presented to the child. According to Yablo, only an appeal to wide content adequately explains how a variety of substantially different pictures, with respect to their intrinsic properties, prompt the same response from the child. He therefore concludes that it is the wide content of the picture, i.e., the fact that it is of the child's mother, which is most relevant to explaining the child's behaviour.

---

[59] The brackets are mine; Yablo describes what is depicted in the photographs as the child's bubbe; I have interpreted this as the child's mother.

Let us now consider whether this example violates any of the methodological constraints of explanation. First, what is the supposed relation that is argued to be causally explanatory? I take this to be the relation between the picture and the child's mother: aRb, where 'a' is the photograph and 'b' is the child's mother, and 'R' is the relation 'picture of'. A distinctive feature of this relation is that one can change many variables associated with 'a', whilst keeping variables associated with 'b' constant. This reverses the account of relations given thus far. It is the changing variables associated with the photograph of an unchanged object that causes a constant behaviour in the child. In the previous examples the variables that are associated with the term which cause the event to be explained are held constant whilst the variables associated with the 'wide' term are changed. In this case, the third event is the behaviour of the child - a happy gurgle. Yablo might be taken to be claiming that it is 'R' which best explains the behaviour of the child and not any variables associated specifically with 'a'; i.e., specific intrinsic properties of the picture.

It is worth noting that one can easily break the relation between the photograph and the mother; perhaps the photograph is in fact of the baby's aunt who happens to be his mother's identical twin sister. Arguably the child will respond in just the same way. How does Yablo's 'picture of' relation explain this? He appears to acknowledge that this is a response by acknowledging that "keeping all the colours alone, and regardless of the subject, Issac grins" (p. 265). But then he immediately claims that:

> Shouldn't we then conclude that Issac's behaviour is controlled more by the picture's intrinsic colour properties than by its extrinsic, representational ones? And if [Isaac's behaviour] [60] is controlled by the colour properties, then the very last thing we would expect is that a *differently* coloured picture of his [mother] would still have lead [sic] Isaac to grin.
>
> And, yet, this is *precisely* what we would expect (p.265).

The most seriously flawed aspect of Yablo's argument is that he commits the intrinsicalist to maintaining that *only* the intrinsic *colour* properties of

---

[60] The brackets are mine.

the photographs are relevant with regards to Issac's behaviour. It is true that focussing on an identity of all the intrinsic properties will result in behavioural identity, and this focus on an identical photograph resulting in the same behaviour ignores the possibly relevant causal explanation. But it is a mistake to assume as Yablo does that once *colour* has been ruled out as the relevant intrinsic property, the wide content of the picture explains Isaac's behaviour. Yablo needs to establish that no shared intrinsic properties between the distinct photographs explain Issac's behaviour. Once he establishes this, he can then argue that only the relational property between the photograph and what it depicts explains Isaac's behaviour. However, there are many plausible shared intrinsic properties, other than colour, that may well be applicable to the wide variety of photographs that all provoke the same response form Isaac.

The points Yablo makes about the intrinsic property colour are plausible, but the damage has been done. He infers that if the intrinsicalist only appeals to the intrinsic *colour* properties of the photograph, then this intrinsicalist position entails that different coloured pictures will change the behaviour of Issac. But why can the intrinsicalist not appeal to other types of intrinsic properties that might be shared by a wide variety of photographic images? Next, he clams that the child does in fact respond to differently coloured pictures the same way. Let us also assume that experimental evidence is consistent with the constancy of behaviour of the child when he is presented with differently coloured pictures. We do expect the child to respond differently to differently coloured pictures. This is the implied *reductio* of the above claim. Yablo's argument can be summarised as follows: the kind of intrinsicalist who keeps on harping on about how identical photographs produce identical behaviour entails the position that different coloured photographs produce different behaviours, but different coloured photographs do not produce different behaviours; therefore the intrinsicalist position is false, and wide contents offer the best explanation.

Through the above reasoning, Yablo commits to someone who dares utter such an obvious truth that a photograph of Isaac's identical aunt will provoke an identical response, the *inability* to handle the more interesting truth that Isaac responds the same way to differently coloured photographs

of the same person. But the intrinsicalist does not need to only appeal to colour as the shared intrinsic properties that a variety of photographs possess that all provoke the same response; therefore, this line of reasoning certainly does not offer a *reductio* on the earlier claim that the photograph of the aunt *does* provoke an identical response in Isaac.

Yablo's focus on colour leads him to entirely ignore the intrinsic properties a variety of photographs and Isaac's mother might share; for example, proportional distances between facial features.[61] The explanation for why the variables associated with the photograph have something in common with the baby's mother is not the *relation*, but some causal story about how variables associated with the mother are explanatorily relevant to variables associated with the photograph or even a set of different photographs of the mother. Does the relation between the photograph, or the set of photographs and the child's mother explain a third event - the child's behaviour? Yablo argues, yes. However, once again what has to be spelt out is how the variables associated with the photograph, or a set of photographs, comes to cause certain responses of the child. This immediately poses the question as to how the child matches variables, associated with a photograph, with something it recognises. Does the relation between the mother and the photograph explain this? Arguably not.

Could there not exist certain variables, associated with all of the photographs of the mother, and incidentally of the mother's identical twin, which all the photographs share, other than merely the *relation* between the photograph and the child's mother? These shared variables associated with the photographs might be small, but they might be important; therefore, according to the anti-wide representationist, it is not in the least *hard to understand* why Isaac should lose sight of his mother in the alternative-image world nearest to this one. Perhaps, as already suggested, the important, but *small* variable that explains why Isaac can respond happily to a variety of photographs is the proportional distances between all the facial features depicted in the photo; i.e., the relative distances between the eyes,

---

[61] Whether this is right is irrelevant; the main point is that one has to do the experiments; one does not rest content with the 'content' of the picture explaining the behaviour of the child.

the ears, the nose and the mouth and the chin.[62] Incidentally, the proportions maintained between features of a face is, somewhat obviously, a shared variable of both the photographs, the child's mother and the child's 'identical' aunt; furthermore, these proportions will be maintained through the causal process of the photograph's construction. Yablo is merely content that one variable, the 'picture of' relation between a photograph, or a set of photographs, and an object best explains the child's behaviour in this situation. However, what about this small but interesting proportionality variable that both the variety of photographs, and the child's mother, share? Furthermore, an emphasis on only the relational property of the photographs ignores the fascinating account of as how children are able to recognise two-dimensional images and match them to familiar objects.

What about the actual perceptual and cognitive abilities of the child? To merely explain the behaviour of the child in terms of the relation between the photograph and the mother ignores the arguably complex and essential story concerning both the perceptual and cognitive abilities of the child. Once again there is a detecting mechanism within the child that might be bringing together variables that initially one would have thought would have no normal statistical relevance. Somehow the child is clearly recognising some variables associated with a variety of photographs and linking them with variables associated with an object, which happens to bear some relation to the photographs. How the child does this is not going to be explained by only the picture relation that a set of pictures stand in, relative to an object that is depicted in these photographs. Any cognitive psychologist would be mad to accept this as an adequate explanation of the child's behaviour.

If an explanation is given as to how a child recognises its mother, by sight, from a variety of different angles, then arguably this explanation will go a long way to explaining why a child responds to a set of distinct pictures in the same way. Notice this explanation makes no reference to relations

---

[62]Vertical proportions are maintained in profiles: forehead to eyebrows to eyes to nose to chin. Horizontal proportions are maintained when faces are looking down or up: ears to eye to nose to eye to ear. This entirely ignores the importance of shapes of features; however, I see no reason why they should be not shared by a whole variety of different coloured photographs.

between photographs and objects depicted in the photographs, although there may be intrinsic properties shared by the photographs and the mother. Maybe, as suggested, proportions between facial features are relevant for facial identification. No doubt there might be variables associated with the shapes of specific features such as colour and style of hair, colour and shape of eyes or colour of skin. Relations between pictures and objects do not explain how children recognise their mothers from a variety of different angles by sight, yet I suspect that whatever explains how a child recognises its mother by sight from a variety of angles, partly explains why a various group of pictures are *all* recognised by the child.

This point about shared intrinsic properties is a small one but it is important given Yablo's attempt to justify that wide contents might be more suitable for causal explanation. He sets the anti-wide representationist up in a way that is hardly conducive to plausibility in his photograph example by committing them to a somewhat lame set of intrinsic properties - colour. By this one move, he then deems it sufficient to ignore the fact that identical photographs produce identical responses, despite the fact that the content relation between the photograph and what it depicts might be broken, yet this will elicit the same response from the child. But most importantly Yablo ignores the possibility that shared intrinsic properties explain the child's behaviour.

Casting the net wider, it is Yablo's claim that beliefs and all intentional mental states likewise also often vastly vary with regards to their neurological properties, yet they often produce the same responses. This is similar to the traditional point the functionalists made long ago concerning the multiple realizability of functional states. Yablo is really just offering the same point in a different guise. The next chapter will consider in more depth the functionalist objections to type-identity theories and whether the Phenomenist can avoid these problems.

5.6 Conclusion

To conclude both this and the previous chapter, a dilemma is presented to the wide representationist: do they either adopt strict relations or loose

relations as explanatorily relevant? If they adopt strict relations, they face serious problems explaining how the various terms that constitute that relation are explanatorily relevant with regards to each other. These problems arise from four explanatory constraints considered in the previous chapter. If the wide representationist adopts a looser conception of relations, they also face problems. How does the wide representationist justify that their loose relations are explanatory in a way that strict relations are not? They will have to appeal to the variables associated with the terms that constitute the relation and show how they do not violate the above explanatory constraints. The looser conception of relations allows for the terms that constitute the relation to be explanatory in a manner that meets the explanatory constraints; however, it is not clear, even on the looser conception, how the relation itself contributes in any way to a *deep* explanation.

The anti-wide representationist need not maintain that the variables associated with the terms that constitute the relation are never explanatorily relevant in the sense that they fail to describe regularities; this would be a straw anti-representationism. The anti-representationist merely maintains that the wide representationist faces a problem in accounting for how these supposedly explanatory relations do not violate some of the explanatory constraints considered in the previous chapter without appealing only to the variables associated with the terms that constitute the relation. In the cases of the biological examples cited, it is clear that the loose relations appealed to do not violate explanatory constraints because of often indirect complex processes that are identified when the variables associated with the terms that constitute the relation are identified and are found to be statistically or counterfactually relevant to other variables. The biological examples also appeal to third party perceptual mechanisms to bring together normally unrelated variables. It is not clear that the wide representationist hopes that wide contents explain by an appeal to third party detecting mechanisms. An appeal to a simple cognitive experiment involving a child responding to a variety of photographic images also does not give adequate grounds for claiming that wide contents are causally explanatory.

In the picture example that Yablo cites, he successfully establishes the point that only claiming that identical pictures produce identical responses hardly offers a deep explanation for those responses. This point will be restated and developed in the next chapter as an attempt to demonstrate that the Phenomenist who wishes to maintain an explanatory role for phenomenal character is required to adopt an identity theory that is problematic on functionalist grounds. However, Yablo's example fails to demonstrate that there are no shared intrinsic properties of the photographs that explain the child's similar response. Granted these problems might arise for Yablo because of the quality of his example; however, it is perhaps revealing how his example fails to establish that extrinsic properties buy up an abundance of intrinsic. More plausible accounts of the relevant shared intrinsic properties might well buy up an abundance of extrinsic.

The phenomenist also confronts problems concerning an explanatory or causal role for phenomenal content or character. These problems will developed in the next chapter.

Chapter 6: Phenomenism and Causation

6.1 Introduction

In chapters four and five, it was demonstrated that the wide representationist fails to adequately offer an account consistent with wide contents having a role to play in deep explanations. Considerations concerning causation motivated Fodor to give his narrow content account where narrow contents are functions that map from contexts onto truth conditions. However, although narrow content accounts help avoid problems concerning causation, in chapter three it was demonstrated that arguably the best narrow content account of phenomenal character fails to adequately account for phenomenal character. Might one conclude that phenomenism is the best theory? Unfortunately not. It is not clear that the phenomenist is any better off with regards to some of the considerations concerning causation and explanation that were developed in the previous chapter.

This chapter presents two dilemmas to the phenomenist as follows: The first gives the phenomenist the choice of adopting either a physicalist or a non-physicalist approach to phenomenal character; it will be argued that if the phenomenist wishes phenomenal character to have a causal role, they should opt for a physicalist approach, but then they are confronted with the 'knowledge argument' (Jackson, 1982). The second dilemma confronts the phenomenist who adopts a physicalist approach. A choice is presented to the phenomenist in terms of the variant of identity theory they wish to hold, either a token identity theory or a type identity theory; however, whichever identity theory they adopt they confront substantial difficulties. If the phenomenist adopts a token identity theory, explanatory and causal concerns are problematic, type-identity theories face problems grounded on concerns motivated out of functionalist objections.

The wide representationist faces problems concerning the causal efficacy of relations. These problems were developed in the previous two chapters partly by the employment of twin-earth style examples. The phenomenist also faces a possible metaphysical speculation, analogous to the twin-earth scenarios, whereby the supposedly relevant variables *change* across two

individuals, but this has absolutely no causal relevance to the behaviour of the two individuals - the inverted spectrum. According to a non-physical variant of phenomenism, an inverted spectrum might be supposed to be possible across two physically identical individuals, thus immediately resulting in deep causal and explanatory problems for phenomenal character.

However, inverted spectrums might also be possible across two physically distinct but functionally isomorphic individuals. If the two individuals are functionally identical but their mental states have different physical realizers, then functionally isomorphic individuals might be physically distinct. According to the physicalist variant of phenomenism, the phenomenist identifies phenomenal characterwith the physical realizers of these functional roles; thus spectum inversions might be possible across functionally isomorphic individuals. According to both physical and non-physical variants of Phenomensim, in inverted spectrum scenarios there is a change in either the physical or the non-physical intrinsic properties across the two individuals, but either way these changes have no relevance to the ensuing behaviour of the organism.

On Twin Earth there is no change in the intrinsic properties of the two individuals, yet the mental states are supposed to be distinct, and the behaviours are arguably identical. When comparing Twin Earth to an inverted spectrum scenario, the non-physicalist variant of phenomenism faces similar problems. Two individuals might be identical with respects to all their intrinsic physical properties, yet they experience different phenomenal characters or contents. But if they have identical intrinsic physical properties, then there is no reason to suppose that their behaviours will be any different; therefore, the differing phenomenal characteror content has no relevance to the ensuing behaviours of the two individuals. This suggests that if a non-physcialist phenomenist employs inverted spectrum scenarios against the wide representationist, they are almost certainly going to face some of the causal challenges the representationist faced in the previous chapter.

The physicalist variant of phenomenism is committed to some kind of identity theory between phenomenal characterand the intrinsic properties of

brains. However, this physicalist variant of phenomenism might be presented with a further challenge for two reasons. First, if the phenomenist opts for a token-identity theory, whereby a *token* phenomenal character is identical to *a token* intrinsic physical property perhaps in addition to a limited supervenience claim,[63] then token-identity, if combined with mere supervenience between phenomenal character and intrinsic properties, is not sufficient for an account of intrinsic properties being deeply explanatory.[64] Token-identity and mere supervenience offer a limited claim that leaves no room for what might be described as explanatory 'power'.[65] Second, if the phenomenist opts for a type-type identity theory, where types of intrinsic properties are identical to types of phenomenal character, then the phenomenist gains explanatory 'power' for intrinsic properties, but at the cost of well rehearsed functionalist objections to type-identity theories of mental states.

This chapter further explores some of the above considerations thus presenting causal and explanatory challenges to both physicalist and non-physicalist variants of phenomenism.

6.2.1 Explanatory gaps

The first dilemma that confronts the phenomenist is motivated by what has sometimes been described as an 'explanatory gap' (Levine, 1993). It is perhaps appropriate that the phenomenist should also face objections

---

[63]There might be some debate concerning what is sometimes described as the 'width' of the supervenience claim. The wide representationist is consistent with one type of supervenience, a more 'global' supervenience. The phenomenist and the narrow content representationists maintain that the mental 'locally' supervenes on the intrinsic physical properties of an individual.

[64] This variant of phenomenism is *perhaps* compatible with a very loose formulation of functionalism. The functional role does not constitute the phenomenal character; instead the intrinsic physical properties that realize the functional state constitute the phenomenal character.

[65] It might be argued that a physicalist is just concerned with demonstrating how a property fits into the world of causes as outlined by the natural sciences. It is, however, establishing a deep explanatory role for these properties that this type of phenomenism is trying to establish for phenomenal character or content. Arguably a form of functionalism that holds the realizers satisfies a weak criteria of physcialism but it is problems confronting how second order functional roles could have a causal role that motivates an objection to looser conceptions of physicalism. See Kim (1993).

grounded on metaphysical speculation. The phenomenist faces an objection based on a 'knowledge argument' that arguably establishes this troubling 'explanatory gap' between phenomenal character and the intrinsic physical properties that the phenomenist might take to constitute phenomenal character(Jackson, 1982, Levine, 1993). This argument, if successful, is sufficient for establishing both the non-physicality of phenomenal character and hence the deep epiphenomenal nature of phenomenal character.

Frank Jackson (1982) holds that phenomenal characteris both non-physical and epiphenomenal. His position somewhat starkly illustrates that if the phenomenist holds that phenomenal characteris not physical, they are going to face considerable problems accounting for a deep explanatory or causal role for phenomenal character. Jackson, however, embraces a non-causal or non-explanatory role for phenomenal character. Jackson proposed his 'knowledge' argument as an objection to any physicalist account:

> Nothing you could tell of a physical sort captures the smell of a rose, for instance. Therefore, Physicalism is false (p.127).

Acccording to Jackson, this failure of physicalism entails that phenomenal character is epiphenomenal. He demonstrates this 'explanatory gap' and supports the above brief argument by the use of a metaphysical speculation - Mary's Room.

Mary is a scientist who only perceives in black and white. She studies colour in a room where there is no colour other than black and white and her only means of seeing the world is through a black and white television monitor. The intuition is that Mary could learn everything of a third-person objective sort that there is to learn about colour, she could know all the relevant facts about the intrinsic physical properties and the physical processes, but given all this knowledge, she would never know what it was really like to experiencered. When she sees red for the first time she learns something *new*.

> What will happen when Mary is released from her black and white room or is given a colour television monitor? Will she *learn* anything or not? It seems just obvious that she will learn something about the world and our

> visual experience of it. But then it is inescapable that her previous knowledge was incomplete. But she had *all* the physical information. *Ergo* there is more to have than that, and physicalism is false (p.129).

If this argument is successful and there genuinely is some kind of explanatory gap between phenomenal character and all intrinsic physical properties, then the phenomenist is left in the deeply unpalatable situation of embracing the absence of any explanatory role phenomenal character has to play in explaining behaviour. This is in contrast with Jackson who clearly is willing to embrace the epiphenomenality of phenomenal character.[66] I will consider arguably the best line of response to this argument. If this reply fails, then there are at least *prima facie* grounds for maintaining that as the representationist faced problems in giving phenomenal content or character an explanatory role, so does the phenomenist.

6.2.2 Modes of presentation

Arguably, the best rebuttal to Jackson's argument appeals to different modes of presentation - different ways that the same properties might be presented to an individual.[67] For example, I could come to know the score of a

---

[66] "[I]t is possible to hold that certain properties of certain mental states, namely those I've called qualia, are such that their possession or absence makes no difference to the physical world" Jackson, 1982, p.132).

[67] Dennett attempts to rebut Jackson's argument by demonstrating that it is impossible to get clear on the intuition that Mary knows all that there is to be known about the physical facts. Therefore, because of this difficulty, the argument fails to establish even the first premise. An alternative is to simply deny that Mary learns something new. This view can be seen in Churchland (1985), Dennett (1991) and Akins (1993). This reply is however problematic because it blocks the fairly plausible intuition that knowledge of propositions still fails to allow the knowledge of the sort Mary gains when she sees red for the first time (Papineau, forthcoming).
An alternative reply is to argue that Mary gains new abilities, but not new knowledge. This is perhaps analogous to learning the knowledge about how to windsurf, but until one actually learns to windsurf, one does not possess the ability. Mary can predict who experiences red, but until she actually experiences red she does not have the introspective ability to experience red. She does not learn something new; she merely gains a new ability. For two defenders of the ability response see Lewis (1988) and Nemirov (1990). However, if gaining the new ability is sufficient for Mary to gain new knowledge, then after she gains the ability Mary learns something new. She arguably not only gains the ability to introspectively detect whether an object is red without using her technical instruments, but she learns how to correlate being presented with red with her specific experiences. This new means to knowledge, however, gives her a new way

football game by watching the match in person, or I could know the score of the football game by reading it in the newspaper the next day. Either way, I would possess knowledge of the same property but under different modes of presentation. Perhaps when Mary leaves the room for the first time she comes to know something she already knew but under a different mode of presentation (Lycan, 1996).

First, let us consider how Mary could know that an object is red. There are two ways of knowing that an object is red. Mary could know this either by an examination of the reflectance properties, or she could know this by using her own perceptual capacities. What Mary lacks is not knowledge, but knowing under a different mode of presentation. It is not surprising that Mary lacks the mode of knowledge of detecting some object to be red through experience because she has never detected a red object through directly perceiving it. Mary does not learn something new when she comes to perceive red for the first time; she merely comes to know something that she already knew in a different way.

But if Mary knew everything, surely she should know what it is like to know something under a different mode of presentation. Why can she not infer what it would be like to know a football score through reading it in a newspaper or seeing the game in person? Should Mary not already know what it is like to see red for the first time under a different mode of presentation if she genuinely knows all the relevant facts?

Papineau (forthcoming) and Tye (1995b) suggest that Mary does not possess the phenomenal concepts necessary to know that an object is red under the mode of presentation that constitutes perceiving the object 'as red'. The theoretical description she does know involves no phenomenal concepts, only material concepts; therefore, Mary cannot infer what it would be like to know the same fact under a different mode of presentation. Essential to this line of response is that you cannot, from the entire

---

to conceptualise the knowledge that an object is red she can come to think the new thought that she is having a phenomenal thought of this type (Tye, 1999). This presumably gives her new knowledge. For more arguments against the ability response see Lycan (1996).

knowledge content of one mode of presentation, infer what it would be like to know the same knowledge under a different mode of presentation. There is a 'conceptual gap'.

Thus far the reply to Jackson has been run in a way that assumes that phenomenal character is entirely constituted by intentional content and Tye (1995b) think that this helps him formulate his response to Jackson, but as Levine (1997, p.105) argues, it is not clear that the phenomenist cannot also run a similar defence just as successfully against Jackson's argument.

If the knowledge that Mary is supposed to know is about the intrinsic physical properties that are supposed by the phenomenist to constitute the phenomenal character of an experience, a similar argument can be run as a reply to Jackson. Mary knows under only one mode of presentation as to how intrinsic physical properties are type identified with phenomenal character. She could therefore know if a person were to experience a specific phenomenal character of a certain type - including she, just as she could know if an object was in fact red. However, it so happens she has never come to posses the alternative introspective mode of presentation of herself experiencing red because her mental states have never been in that physical state (Papineau, forthcoming). She has not come to know what it is like to experience red because she has never previously possessed the necessary phenomenal concept (Tye, 1999).

6.2.3 Modes of presentation fail to offer an adequate reply

Joseph Levine (1997), however, thinks that any attempt to appeal to an introspective mode of presentation as a response to Jackson's 'knowledge argument' has serious problems. If Levine is right, then arguably the best response to the knowledge argument fails and the phenomenist is still left with a troublesome explanatory gap. He argues that there always remains the question as to why one mode of presentation cannot be used to infer what it would be like to know that same knowledge under a different mode of presentation. He thinks that the only meaningful way to answer this question is to allow that the introspective mode of presentation in the case of qualia ascribes new properties, but this then entails that Mary really does

learn about some new properties over and above the physical properties that she already knew about. There is an alternative way of differentiating between modes of presentation and that is to appeal to a demonstrative or indexical mode of presentation. One might know lots about coffee cups, but one can also learn that *this* a coffee cup by pointing at it, thus applying the knowledge to a specific cup. But according to Levine, demonstrative modes of presentation fail to have any real cognitive content so they fail to account for the explanatory gap. The only way to generate a gap that allows for one not to infer one mode of presentation from another is to allow that they ascribe different properties, but this then entails that all the intrinsic physical properties under one mode of presentation fail to capture all the properties that are available to be known.

According to the mode of presentation reply to the knowledge argument, Mary does not possess the alternative mode of knowledge before she sees red for the first time. But she does not undergo the new mode of presentation, before she sees red, not because there is some property of mental states that she does not know about. She knows about all the intrinsic physical properties of mental states. To say that it is a phenomenal property over and above the physical property that explains the failure of a different mode of presentation is to concede to Jackson's point that phenomenal properties are properties over and above intrinsic physical properties. But the only way that one can account for an individual not having access to a distinct mode of presentation is because there is something about the intrinsic physical properties that precludes access to a different mode. Therefore the modes of presentation reply to Jackson's argument fails (Levine, 1997).

Levine illustrates this argument by replying to two examples a proponent of this reply to Jackson might draw upon. The first is an example of the morning and the evening star. There is clearly a sense, given an ignorance of astronomy, that one would have a knowledge of the star under two distinct modes of presentation. Furthermore it is clear that one cannot infer that the star is the same under the distinct modes of presentation. But the reason why is because there are differences in the properties that you ascribe to the star, one appears in the morning and the other appears in the evening.

> [T]he difference consists in a difference in the properties ascribed: appearing in the morning in one case and appearing in the evening in the other. So this case can't serve as a model for the explanatory gap, since the materialist opponent of the gap doesn't want to concede that the introspective mode of presentation differs from the third-person mode with respect to the property of the qualitative state by which it represents it. For this would mean that the qualitative character is not the same property as whatever physical or functional property is characterised by the third-person mode, and we are then left to wonder how the latter property gives rise to the former (Levine, 1997, p.107).

In this example it is fairly clear that a difference of properties ascribed to the star explains why an individual does not infer that the star that appears in the morning is the same star as the star in the evening. But, as Levine points out, it is precisely this appeal to distinct property ascriptions to justify the distinct modes of presentation that the person relying on distinct mode of presentations cannot appeal to in order to reply to the knowledge argument.

However, the proponent of the mode of presentations reply to Jackson's argument might adopt another example to illustrate how modes of knowledge might be distinct without an appeal to a difference in the properties that constitute the knowledge - the messy shopper.[68]

The messy shopper is someone who walks around a shop following a trail of sugar. He eventually discovers that the sugar is leaking from his own cart. The point of this example is that one could know all the relevant facts about trolleys, shops, and leaking sugar, but one would still fail to know that it was oneself that was leaking the sugar. But when one does realise that it is oneself who is leaking the sugar, one does not learn anything new; one merely applies the knowledge to oneself. One gains a demonstrative form of knowledge by coming to see that the messy shopper is in fact you. Furthermore, one cannot infer from all the theoretical knowledge that it is oneself that is leaking the sugar if one does not realise that you are not the shopper.

---

[68]Levine borrows this example from Perry (1979).

This might then be applied as a response to Mary's room. No matter how much she knows theoretically about phenomenal characterit is not until she actually experiences *this* shade of red, through an introspective demonstrative mode of presentation, that she understands what it is like to experience something under a different mode of presentation. Just as the messy shopper goes "aha, the shopper is me!" Mary exclaims: "aha, this is what it is like to experiencered!" Neither individual learns something new; they just demonstratively apply the knowledge to a specific example.

However, Levine argues that the second example also fails to demonstrate that one can have different modes of presentation in the phenomenal case without appealing to a difference in the properties that form the content of that knowledge.

> In the qualia case, the apparent incommensurablility of the two modes of presentation produce a genuine, perfectly intelligible question concerning how or why the two modes of presentation apply to the same thing. In the messy shopper case there is no corresponding question. I don't wonder how *I* could be the messy shopper, once I learned that my grocery cart is the source of the leak (p.108).

Levine goes on to claim that there is no real cognitive content to demonstrative claims about experience, whilst in the case of experience there is clearly a new property ascribed in a manner disanalogous to demonstratively applying knowledge to one instance.

> There is no genuine property of 'me-ness' or 'thisness' to contrast with genuine properties like leaking sugar . . . Therefore there is no genuine cognitive content to questions like 'how is it *me* that is leaking the sugar'? . . . However, because being reddish or greenish is a genuine property, and my introspective mode of presentation is a property-ascribing mode, there is a perfectly contentful intelligent question to ask: what explains how a state satisfying the relevant theoretical description in physical or functional terms also instantiates being reddish (p.108)?

It is presumably the answer to this question that Mary should know given her extensive knowledge, but seeing as she does not, phenomenal properties are properties over and above intrinsic physical properties.

This analysis of Levine's attempt to salvage the knowledge argument against arguably the most plausible way out of the knowledge argument on behalf of the phenomenist hardly offers a solid ground for the existence of an explanatory gap but it does offer perhaps *prima facie* grounds for thinking that the phenomenist has got a considerable burden to demonstrate as to why Mary does not learn something new given an extensive knowledge of all the intrinsic physical properties. The knowledge argument is not the only grounds for maintaining that there is an explanatory gap between phenomenal characterand physical properties.

6.3.1 Inverted spectrums and causal constraints once again

In the inverted spectrum scenario two individuals are completely functionally isomorphic, yet they undergo mental states with distinct phenomenal character. The phenomenist has a choice in accounting for the differencein phenomenal characteracross the two individuals in the inverted spectrum scenario as follows: first, to appeal to the *physical* intrinsic properties of the two individuals being different, hence the phenomenal charactersare different across the two individuals, but this has no impact on the functional roles of the mental states; or second, to appeal to *non-physical* properties, so the two individuals might be identical with regards to every intrinsic physical property, but they are different with regards to intrinsic non-physical properties.

If the phenomenist appeals to an identity between phenomenal characterand non-physical properties, then the phenomenist certainly faces considerable challenges accounting for causation between mental states and physical states. There are no differencesin any of the intrinsic physical properties of the two individuals so there is no reason for maintaining that their behaviours will be different and there is certainly no reason for maintaining that the phenomenal character of the two individuals will have a deep explanatory role to play concerning the behaviours of the two individuals. The wide representationist faces subtle explanatory problems; however, the

non-physicalist variant of phenomenism faces obvious deep explanatory problems.[69]

6.3.2 Nomological laws, non-physicalist phenomenism and inverted spectrums

If one assumes some kind of functional law-like correlation between variables associated with phenomenal character, which according to the phenomenist are identified with the intrinsic non-physical properties of an individual, and variables associated with the behaviour of a creature, then this correlation might take the form of a nomological law that one might argue is required to meet the deductive nomological constraint on explanation. However, in inverted spectrum scenarios there are changes in one of the variables of the above correlation, yet the other variables that form the other part of the above correlation necessarily do not change. This is because, according to the phenomenist, the intrinsic non-physical properties of one individual are *different* across two individuals, yet variables associated with both the intrinsic physical properties of the two individuals and the behaviours of the two individuals are the *same*. The phenomenal characterof one individual's mental states whilst he looks at grass is 'green.' The phenomenal characterof the other individual's mental states is 'red' whilst he looks at grass. However, the intrinsic physical properties and the ensuing behaviour of the two individuals is the same.

If there is change in one variable of a correlation whilst there is necessarily *no* change in another variable of the correlation, it is not clear how a correlation between the variables can possibly be constructed. If no correlations can be established between phenomenal characterand behaviour, then no nomological laws can be constructed between phenomenal character and behaviour. Therefore, on a deductive nomological conception of explanation that relies on the presence of nomological laws, phenomenal

---

[69] It might be suggested that differences in phenomenal character do the causal work (my examiners raised this point). This would give a causal role to phenomenal character, but notice that it would once again be relations that would be doing the causal work and not the intrinsic phenomenal properties that the phenomenist hopes will bear the explanatory burden.

characterhas no deep explanatory role to play with regards to explaining behaviour.

6.3.3 Statistical relevance, non-physicalist phenomenism and inverted spectrums

The statistical relevanceconstraint on explanation requires that there at least be some kind of statistical relevance between variables associated with events, for changes in the events to be explanatory with regards to one another. In chapterfour it was discussed how relations, on one conception, requirethat one of the variables does not change whilst the other one does; thus making it difficult to see how variables associated with one of the terms has any statistical relevanceto variables associated with the other term (other than the relational property).

According to the non-physicalist variant of phenomenism, in the inverted spectrum scenario there is a change in phenomenal characteracross the two individuals, yet there is no change in the intrinsic physical properties and hence the ensuing behaviours of the two individuals. It appears that variables associated with phenomenal characterhave no statistical relevance at least in this one instance with regards to variables associated with the ensuing behaviour of the organism, nor is it clear how changes in non-physical properties could have any statistical relevance.

6.3.4 Counterfactual relevance, non-physicalist phenomenism and inverted spectrums

Crude Counterfactualism states that event A is *not* the cause of another event B if there exists a nearby possible world where A fails to occur, yet, B nonetheless still occurs.

Inverted spectrum scenarios are set up in such a way that if variables associated with the non-physical phenomenal character of an individual's mental states are different across two individuals, then this difference results in absolutely no change in the intrinsic physical properties of an individual and the behaviour across the two individuals. However, this applies only to

two physically identical individuals who are pre-supposed to behave in the same way within the same possible world. The non-physicalist phenomenist could still maintain a counter-factual causal account for the phenomenal experiences of an individual given a close possible world whereby some phenomenally distinct individual resides who is physically identical, but behaviourally distinct.

In order to present a counterfactual account of causation as an objection to the non-physicalist phenomenist, what is required is the demonstration that in *all* nearby physically identical possible worlds where the phenomenally distinct individual lives, the subsequent behaviour of this individual is *necessarily* identical. However, the non-physicalist phenomenist only requires a difference in behaviours of these two individuals (despite physical identity) in *some* possible world. To merely assume otherwise begs the question against the non-physicalist phenomenist.

The phenomenist could still maintain that there is a possible world where a physically identical individual resides and were that individual to not have that same non-physical phenomenal experiences, that individual would not behave in the same way. I suppose at this point we have to concede the possibility of the non-physicalist phenomenist consistently maintaining a counterfactual account of causation, but is this really plausible?

Do the phenomenal states of the individual in the nearby possible world somehow bypass the intrinsic physical properties of the indiviual at t_1 before the subsequent behaviour of the individual at t_2? If there is any physical difference in the two individuals at t_1, then physiological identity is no longer maintained across the distinct possible worlds and we cannot assume it is the phenomenal difference that accounts for the difference in behaviour.[70] Furthermore, even if the non-physicalist can sustain a counterfactual causal account there remains substantial problems concerning the causal mechanisms of interaction.

---

[70] The indeterminism of the universe might be sufficient for a difference in behaviour across the possible worlds, but the non-physicalist phenomenist requires the difference in behaviour to not be a result of indeterminacy, but the phenomenal difference between the two individuals as the *cause*.

6.3.5 Interactionism, non-physicalist phenomenism and inverted spectrums

Interactionism requires that there be local interactions that mediate between events for the events to be deemed causally efficacious. One might argue that there has to be some kind of exchange of energy for the interaction to be a 'genuine' causal interaction. This is arguably where an appeal to non-physical properties presents even more substantial difficulties for presenting a causal or explanatory role for the phenomenal.

It might be assumed that the non-physical properties I have been attributing to the phenomenist is in fact compatible with some kind of physicalist account, perhaps consistent with what has sometimes been described as property dualism.[71] I will consider this to be a variant of physicalist phenomenism that will be considered in the next sections. The non-physicalist phenomenist under discussion genuinely holds that phenomenal characteris substantively non-physical. The individuals who are spectrum inverted relative to one another genuinely possess substantive non-physicality responsible for their differences in phenomenal characterdespite molecular identity. It is simply not clear how any such phenomenal charactercould have an interactive role.

6.4 Summing up causal constraints against the non-physicalist phenomenist

It appears that the very argument that poses a problem to the wide representationist presents deep explanatory concerns to the non-physicalist phenomenist. The epiphenomenal nature of a phenomenist who rejects claims that phenomenal charactercan be explained in terms of the physical is perhaps not disputable. Jackson does not think that this epiphenomenality is a problem. However, for any of the arguments against the wide representationist that were considered in the previous chapters to

---

[71] Property dualism has the benefit of accounting for how mental states can enter into causal roles, but there remains the problem of whether they can ever play a deep explanatory role. Furthermore, property dualism face a problem of causal overdetermination (Kim, 1993).

have any substantive weight, the phenomenist must maintain some kind of causal or explanatory role for phenomenal character.

The physicalist variant of phenomenism attempts to identify phenomenal character with intrinsic physical properties. The above arguments were partly motivated against any physicalist account of phenomenal character. It appears that, at least with regards to nomological laws, statistical and counterfactual relevance and interaction, the physicalist phenomenist has an advantage over both the non-physicalist phenomenist and the wide representationist. However, in the cases of inverted spectrums that rely on differences in the intrinsic physical properties that realize the same functions of the mental states, it is not clear that the phenomenist who identifies phenomenal character with intrinsic physical properties is necessarily successful in giving an explanatory role for phenomenal character.

The affirmation of *mere* supervenience between phenomenal character and intrinsic physical properties is not sufficient to establish a deep explanatory role for phenomenal character(Kim, p.190, 1997). It is arguably this deep explanatory role for phenomenal character that the phenomenist needs to establish in order to claim that their account is superior to the wide representationist. A dilemma can be presented to the physicalist phenomenist as follows: either they hold a variant on a *token*-identity theory between phenomenal character and intrinsic properties of an individual's mental states combined with the affirmation of *mere* supervenience between phenomenal character and intrinsic physical properties; or they maintain a *type*-identity theory between types of phenomenal character and types of intrinsic properties. If the phenomenist adopts either option, then problems arise.

6.5.1 Token-identity, supervenience and phenomenism

If the phenomenist opts for a variant on token-identity between phenomenal character and intrinsic properties, combined with *mere* supervenience, then a

deep explanatory role for phenomenal characteris lost. [72] The phenomenist arguably requires more. Consider the following claim:

P1 People intrinsically, just like you, when they are presented with a ripe tomato, experience the *same* phenomenal character; and, if they want to describe its colour, say: "look at the red tomato."

This claim is consistent with a token-identity theory of phenomenal charactercombined with mere supervenience. One can only infer from P1 that if someone is intrinsically identical to you, they will have the same phenomenal characterand will behave the same way. Crucially, however, P1 does not explain why the person says: "look at the red tomato," in terms of the phenomenal characterof the individual's mental states. Notice that P1 does some work in Block's presentation of the Inverted Earth argument; helping us to infer that the phenomenal characterof the individual's mental states do not change.[73] Consider this next claim:

P2 People in intrinsic state of type P, when presented with a ripe tomato, experiencea phenomenal character; and, if they want to describe its colour, *ceteris paribus* say: "look at the red tomato."

Where P is some shareable aspect of the intrinsic properties of an individual's mental states. P2 is still consistent with a token-identity theory where a specific type of intrinsic property is correlated to a token phenomenal character. Notice that this claim, P2, cannot be inferred from P1. P1 has plausibility, but fails to give a deep explanatory role for phenomenal character. Even P2 is ambiguous between two further claims. P2 is consistent with an individual having a token phenomenal characterof *some type or other* supervening on an intrinsic property *of type* 'P' that results in the individual exclaiming "look at the red tomato." Arguably the

---

[72]I have drawn on Yablo's argument, where he demonstrates that the Individualist also faces explanatory concerns; however, he runs the above propositions in terms of beliefs about water and water drinking (Yablo, 1997, pp. 290-1).

[73]If the lenses are part of the environment, there is a sense in which the individual's mental states still function correctly (cf with bubbled scenario). If the visual processing is tampered with at a deeper layer, then we can conclude that the Earthling substantially misrepresents all his colour experiences.

phenomenist who wishes phenomenal characterto have a deep explanatory role, with regards to behaviour, requires that all individuals who exclaim "look at the red tomato" not only *share* a type of intrinsic property, but also *share* a type of phenomenal characterthat causes the exclamation: "look at the red tomato"

This is a somewhat elaborate way of claiming that if we want to explain behaviours, such as describing the colours of objects by appealing to the intrinsic properties of an individuals mental states, then only claiming: "if two individuals who are physically identical, then they will behave the same way," does little to offer a deep explanation. As Kim (1997) concludes:

> [M]ind-body supervenience itself is not an *explanatory theory*, it merely states a pattern of property covariation between the mental and the physical, and points to the existence of a dependency relation between the two (p.190).

An *explanation* might be offered by claiming that 'if two individuals share intrinsic properties of a certain type, then they will behave in the same way, and this type of intrinsic property shared by the two individuals explains why they acted in the way that they did.' Token-identities, combined with mere supervenience, are neither sufficient for offering a deep explanation in terms of intrinsic properties, nor are they of any use as premises in an argument, which might be presented, to support the real kinds of identities the phenomenist requires; i.e., type-type identity between phenomenal characterand intrinsic properties.[74] If the phenomenist wants phenomenal characterto have a deep explanatory role, then it is only the shared type of intrinsic properties that explains shared behaviour.[75] Token-identities combined with mere *local* supervenience might serve to rule out wide

---

[74] This is a strong claim that attempts to rule out any multiply realized functional role as being deeply explanatory, but it is presumably exactly this claim that Kim's objections to second order functional roles attempts to undermine on the grounds of overdetermination (1993). It does appear to allow that these functional roles explain, but the explain only in virtue of the realizers that realize those functional roles.

[75]Notice that inverted spectrums violate an appeal to shared type-identity as an attempt to explain behaviour because according to the inverted spectrum scenario, two individuals could have differing types of phenomenal character and hence differing types of intrinsic properties, but this is of no relevance to the ensuing behaviour.

representationism; but token-identity with mere supervenience does little to establish a deep explanatory role for phenomenal character. Type-identity theories offer a deep explanatory role, but type-identity theories face a formidable problem: they appear to be overwhelmingly improbable!

6.5.2.1 Type-identity, phenomenism and multiple-realizability

There is arguably no good reason to maintain that the same types of behaviour are going to be explained by some shared type of intrinsic properties across individuals. Why even suppose that intrinsic types of physical properties are going to be shared across species and even aliens? This point still resonates loudly in the literature against the type-identity theorist. As Hilary Putnam (1975c) argues:

> Consider what the brain state theorist has to do to make good his claims. He has to specify a physical-chemical state such that *any* organism (not just a mammal) is in pain if and only if (a) It possesses a brain of a suitable physical-chemical structure; and (b) its brain is in that psycho-chemical state. This means that the physical-chemical state must be a possible state of a mammalian brain, a reptilian brain, a mollusc's brain (octopuses are mollesca, and certainly feel pain), etc. At the same time, it must *not* be a possible (physically possible) state of the brain of any physically possible creature that cannot feel pain. Even if such a state can be found, it must be nomologically certain that it will also be a state of the brain of any extra terrestrial life that may be found that will be capable of feeling pain before we can even entertain the supposition that it may *be* pain (p.436).

Putnam (1975c), although willing to claim that an identification of such type-identities between mental states and intrinsic physical properties is possible, thinks it highly unlikely:

> [I]f we can find even one psychological predicate which can clearly be applied to both a mammal and an octupus (say 'hungry'), but whose physical-chemical 'correlate' is different in the two cases, the brain-state theory has collapsed. It seems to me over-whelmingly probable that we can do this (p.437).

By 'multiply realized' it is meant that the same psychological states can be implemented by a wide variety of physical states. If the phenomenist maintains that phenomenal character is type identical to physio-chemical

states, then psychological states are not multiply realizable. However, it is 'over-whelmingly probable' that mental states, and presumably the accompanying phenomenal character,are multiply realized. Therefore, these theories of phenomenal characterare highly *im*probable.

Putnam's alternative is not to identify mental states with physical kinds: but instead to identify mental states with functional kinds. But, then, if we are to assume that only *physical* realizations implement these functional kinds, it appears that we need to return to a mere supervenience relation between intrinsic physical properties and the phenomenal. The phenomenist is caught in a loop between maintaining a token-identity theory and a physicalist functionalism combined with a *mere* statement of supervenience that implies no real explanatory power or if there is explanatory power this only arises from the properties of the states that realize that functional role, and an account that maintains a type identification between mental states and physical states that explains the supervenience relation, but cannot account for the truth of multiple realizability.

6.5.2.2 Escaping the loop

One way of avoiding this dilemma, for the phenomenist, is to argue that mere supervenience of the phenomenal on the physical, combined with functionalism, *is* sufficient for some kind of explanatory role for the phenomenal. But if the phenomenist wants these functions to be causal, he is caught on the horns of another dilemma: either to identify these functional states with a heterogenous disjunction of intrinsic physical properties that realize that functional states, or to argue that these functional states are *irreducible* to underlying physical states.

If the functionalist opts for a *reductive* functionalism, whereby the functional state is nomically equivalent to a heterogeneous disjunction of physical realizers, then this functionalist faces the problem of explaining how heterogeneous disjunctive properties can possibly enter into causal explanations. But why are disjunctive properties unsuitable for explanation?

One might suggest that such disjunctions are unsuitable for explanation because they are infinite. But why argue that infinite disjunctions are problematic for reduction (Block, 1997, p.110), and hence explanation? Furthermore, why assume that functional properties are nomically equivalent to *infinite* disjunctions (Block, 1997, pp.110-1)?

Arguably the real reason why heterogeneous disjunctions make bad terms in explanations is because heterogeneous disjunctions are not kinds (Fodor, 1974). They are not kinds because heterogeneous disjunctions are not projectable (Kim, 1992). By this it is meant that one cannot make future projections about heterogeneous disjunctions because the confirming evidence for one of the disjuncts need not have any relevance to the other disjunct. Block (1997) develops the claim that heterogenous disjunctions are unprojectable by the use of an example:

> Consider the putative law that people who have arthritis are helped by Ibruprofen. . . Now pick another disease at random, say Lupus, and consider the putative law that people who have either arthritis or lupus are helped by Ibruprofen. This putative law has a heterogeneous disjunctive property (arthritis or lupus) in its antecedent. Does that prevent it from being well confirmed by the same data that confirmed the original law? Each person who has arthritis also has either arthritis or lupus. So *if* the disjunction is projectable, then each datum that confirms the original law also confirms the disjunctive law. Now we have the principle that if P is "well" confirmed and P entails Q, then the evidence that "well" confirms P also confirms (though perhaps not well) Q. But the claim that people who have arthritis or lupus is equivalent to the conjunction of the following two laws:
>
> People who have arthritis are helped by Ibruprofen
> People who have Lupus are helped by Ibroprofen
>
> And if the law with the disjunctive antecedent is well confirmed by the data that confirms the first of these laws, then the second is confirmed by the same data. But we have no information about lupus in this data base at all, so the assumption that the disjunction is projectable leads to a ridiculous result. Conclusion: the heterogeneous disjunction is unprojectable (pp.113-4).

Given the prior assumption that properties that are not projectable are not kinds, and if a property is not a kind, it is not going to be suitable as a term in an explanation, then the above argument successfully demonstrates at least a *prima facie* concern with including heterogenous disjunctions in

explanations. Therefore, the functionalist who argues that functional properties are nomically equivalent to heterogeneous disjunctions faces a problem concerning how these functional properties can be explanatory. But can the functionalist somehow reject that functional properties are to be identified with a heterogeneous set of disjunctive realizers? As Putnam (1975c) once asserted:

> Granted . . . the brain-state theorist can save himself by *ad hoc* assumptions (e.g. defining the disjunction of two states to be a single 'physical-chemical' state), but this does not have to be taken seriously (p.437).

The cost of not taking the identity of the functional state with the disjunction seriously is that the non-reductivist functionalist faces a problem explaining how functional properties are causally efficacious (Kim, 1993). It is surely only in virtue of the physical states that realize the functional properties that the functional states enter into causal relations with other properties or events. If it is assumed that the functional state M is the cause, in addition to the realized state P, then there is a problem in that too *many* causes are candidates for a third distinct effect P*. This is an expression of Kim's causal exclusion problem.

> Suppose . . . that mental property M is causally efficacious with respect to physical property P*, and in particular that a given instance of M causes a given instance of P*. Given the Physical Realization Thesis, this instance of M is there because it is realized by a physical property, say P. . . .What reason is there for not taking P as the cause of P*, bypassing M and treating it as an epiphenomenon?. . . [I]f we insist on M as a cause of P*, we run afoul of . . . "the problem of causal explanatory exclusion." For we would be allowing two distinct sufficient causes, simultaneous with each other, of a single event. This makes the situation look like one of overdetermination, which is absurd. And *ex hypothesi*, it is not possible to regard M and P as forming a single jointly sufficinet cause, each being individually necessary but insuficient. And given the assumed irreducibility of M, we cannot regard M as identical with P, or as part of it. The exclusion problem, then, is this: Given that P is a sufficient cause of P*, how could M *also* be a cause of P*? What causal work is left over for M, or any other mental property to do? M's claim as a cause will be weakened especially if, as we would expect in real-life neurobiological research, there is a continuous causal chain, a mechanism, connecting P with P*. It is clear that the exclusion problem cannot by resolved within the

> framework of nonreductive physicalism (Kim, 1993, pp. 354-5).

The usual reply from the non-reductive physicalists is to embrace some kind of causal compatibilism; whereby high-level properties are deemed to not be reducible to any low level properties, yet are still deemed to be causally efficacious. In the last chapter we examined similar attempts, by wide representationists, to appeal to the explanatory power of high-level relational properties, independent of the relevance of the variables associated with the terms that constitute the relation.

One possible example that the non-reductive functionalist might give is the generalization: "dormative substances, if ingested before driving, cause traffic accidents" (Antony & Levine, 1997, pp.92-3). According to Antony and Levine this generalization is 'realization independent.' It really does not matter how the dormativity is realized in the individual, this dormative property results in the greater chance of a driving accident (p.93).

However, the specific type of dormativity induced by the drug might well have some statistical relevance on the type of accident. Arguably marijuana results in a very different driving behaviour than alcohol, which in turn results in different types of driving accidents, yet both induce dormativity. Once one identifies the variables of a very loose generalization more carefully, it quickly becomes clear that the specific effects of dormative drugs might well be relevant to specific types of driving behaviours, and hence different types of driving accidents. It is only when very loose generalisations are considered that one can hedge the realizations. Furthermore, if two distinct types of drugs cause the very same types of behaviour in an individual, then perhaps there is some property that these drugs share that result in the identical behaviours of the two individuals.

6.5.2.3 Limiting the realizations

It appears that the best line of argument for the phenomenist is to identify phenomenal character with intrinsic properties, accept the possibility of multiple realizability, but attempt explanatory damage control by limiting the level of multiple realizability. Multiple realizability might be limited

by demonstrating that if there are multiple realizations, the physical realizalizations that constitute the disjunction are not as heterogeneous as the functionalist often likes to suggest (Block, 1997).

Kim (1997) argues that if you consider just one *class of system*, fix the physical constitution of that system, and fix the laws of nature, this *fixes* the intrinsic properties that realizes a function in that system.

> Although the realization relation can shift . . . it is also important to note its constancy: Once the systems physical constitution and the prevailing laws of nature are fixed, that fixes whether or not P realizes M in that system. That is to say, if P realizes M in system s, then P will realize M in all systems that are subject to the same laws and are relevantly similar to s - that is, in respect of all nomic properties. If, as most of us would accept, the microstructure of a system determines its causal/nomic properties, it follows that, with laws held constant, the realization relation remains invariant for systems with similar micro-structures.
>
> Consider a class S of systems sharing a relevantly similar microstructure. Biological conspecifics may constitute such a class.
>
> [T]his means that the physical realization provides an *explanation* of mind-body supervenience. The mental supervenes on the physical because every mental property is a second order functional property with physical realizers. And we have an explanation of mental-physical correlations: Why is it the case that whenever P is instantiated in a system, s, it also instantiates mental property M? Because having M consists in having a property with causal specification D, and, in systems like s, then, having M *consists in* having P. It isn't that when certain systems instantiate P. mental property M magically emerges or supervenes, and that this psychophysical correlation must be taken as a brute unexplainable fact. It is rather that having M, for these systems, *is* just having P, or P is one of the ways of having M. This must, by any reasonable standards, be sufficient to warrant the reductive claim that having M, for these systems, is "nothing over and above" having one of its physical realizers(p.197).

By limiting the class of systems considered, one might consider only the human species as the system under consideration (Kim, 1992); this arguably vastly limits the possible realizations a specific mental state can take.[76]

---

[76] As Terrence Horgan (1997) points out, notice how close this is to Armstrong's (1984) solution: "It may be granted for the reasons just discussed, that it is implausible to identify the type *pain* with a certain

However specifying the class of system under consideration raises a rather perplexing question: how large a class of systems should one consider whilst still maintaining a narrow enough class whereby the realizers are limited? Should one consider realizersof mental states in only yourself? In your family? All humans? All primates? All mammals? Any *a priori* speculation on such matters is arguably suspect to doubt. Perhaps there are better ways to limit realizersother than by narrowingthe class of systems to be considered.

Block (1997) suggests limitations on the realizers by arguing that the disjuncts that realize a functional state might not necessarily be as heterogeneous as the functionalist so often suggests. It is easy to get carried away by the vast arrayof possible realizersof a specific functional state. He characterisesthis point with his Disney principle.

> In Walt Disney movies, teacups think and talk, but in the real world, anything that can do those things needs more structure than a teacup. We might call this the Disney principle: that laws of nature impose constraints on ways of making something that satisfies a certain description. There maybe many ways of making such a thing, but not just any old structure will do. It is easy to be mesmerized by the vast variety of different possible realizations of a simple computational structure say that of an *and* gate, which can be made out of cats, mice and cheese as well as mechanical or electrical components. But the vast variety might be cut down to a very few when the function involved is mental, like thinking, for example, and even when there are many realizations, laws of nature may impose impressive constraints. . . . The reductionist may say that cutting down on the possible realizations still allows heterogeneous realizations, but this idea ignores the fact that constraints impose similarities (p.120).

It is the similarities between the realizers of functional states that Block suggests should be emphasised. For example, the fact that two chemicals realize dormativity in an individual suggests not the amazing and vast

---

neurophysiological process. But what about the more narrowly conceived type: *pain in human beings*? It is quite plausible that it can be identified with some single sort of neurophysiological process. And if even that identification turns out to be too optimistic, it will presumably be possible to find still more narrowly conceived sub-types: pain in human beings of the sort Y, . . . and so on, where the identification can finally be effected" (p.162).

heterogeneity of the realizing disjuncts, but that perhaps there is some 'real' property that these two chemicals share (ibid p.127).

6.5.3 Summing up the objection to the physicalist phenomenists

The physicalist phenomenist suggests that phenomenal character should be identified with the intrinsic physical properties of a mental state. But can intrinsic physical properties play an explanatory role with regards to behaviour? Wide representationism was posed a challenge on the grounds that relations are not suitable for deep explanations, but do intrinsic physical properties fare any better when it comes to explaining behaviour? Consideration of causal mechanisms, is certainly problematic to the non-physcialist phenomenist who wishes to maintain the explanatory role of phenomenal character

If the phenomenist opts for the mere supervenience of phenomenal character upon intrinsic physical properties, this falls far short of offering an explanatory role for intrinsic properties. If the phenomenist presents a type-type identity theory, they gain explanatory power but at the cost of considerable implausibility, due to concerns about multiple realizability. The question remains as to whether the phenomenist can steer a course that allows for the phenomenal being identified with intrinsic physical properties, whilst maintaining an explanatory role for these intrinsic properties, whilst allowing for multiple realizability. Non-reductive functionalism entails multiple realizability but is inconsistent with identifying phenomenal character with intrinsic properties. A reductive functionalism allows for the identification of phenomenal character with intrinsic properties, but the resulting heterogeneous disjunction of realizers are *also* not conducive for an explanatory role.

The final suggestion is that the physicalist phenomenist opts for a reductive functionalism that allows for the possibility of limited realizability. The realizing disjuncts that constitute the functional roles are not to be considered as heterogeneous, but largely homogenous. What needs to be emphasised is the commonalities between the properties that realize the disjunction, and not how the realizers differ. This account gives a very clear

explanatory role for the intrinsic properties, both with regards to the fact that they implement a functional role, and how they are part of a shared set of nomologically possible realizersthat through their commonalities explain why they belong to that set, but *not only* because the realizersplay the same functional role.

6.6 Conclusion

This chapter has considered some explanatory problems that face the phenomenist who attempts to appeal to intrinsic physical properties in order to justify how phenomenal character can play an explanatory role in behaviour. The knowledge argument and the non-physicalist variant of inverted spectrum scenarios suggest a strong pro-phenomenist intuition concerning the presence of phenomenal character but this phenomenal characteris of a sort that remains fundamentally irreducible to any intrinsic physical properties. The ensuing epiphenomenalism is arguably untenable for any phenomenist who wishes phenomenal character to have a deep explanatory role. Given that the arguments of the last chapters against the wide representationist were grounded on the assumption of a causal explanatory role for phenomenal character, if the phenomenist embraces epiphenomenalism, then the arguments of the last two chapters that were levelled against the wide representationist are entirely redundant.

However, the physicalist phenomenist also faces problems concerning the necessity of appealing to some kind of type-type identity theory to give phenomenal characterthe desired explanatory role. Mere supervenience only offers a dependency relation but stops well short of a deep explanatory role for phenomenal character. But if the phenomenist adopts the type-type identity theory, the phenomenist has to face the traditional objections from the functionalist theories based on multiple realizability. Some attempts to reconcile the problems of multiple realizability with a plausible identity theory were considered on the grounds of limiting the realizers and appealing to shared intrinsic properties of the realizers. However, even if this reconcilliation can be successfully achieved, the physicalist phenomenist still has to face the problems of the explanatory gap generated by Jackson's knowledge argument.

Chapter 7: Conclusion

7.1 Overview

> How does subjectivity relate to the determinants of content? How should our theory of what an experience is *like* connect with our theory of what it is *of*? . . . The whole question of consciousness and content needs thorough examination (McGinn, 1989, p.63)

This thesis has barely begun to do justice to a thorough examination of the relation between intentional content and the phenomenal content or character of mental states. It has attempted to critically evaluate two approaches to this issue. The first tries to claim that the phenomenal content or character, in Colin McGinn's words 'what an experience is like', is entirely constituted by intentional content or 'what it is of' - this theory is representationism. The second theory, phenomenism, maintains that phenomenal content or character is something more that mere intentional content.

Two general types of considerations constituted the two parts of the thesis. The first part attempted to consider metaphysical speculations to attempt to establish the superiority of one theory over the other. The metaphysical speculations did not, however, conclusively establish either theory to be the best as is perhaps not surprising, but they did serve an invaluable heuristic role in clarifying distinct types of representationism as they attempted to respond to these metaphysical speculations. The second part of the thesis evaluated the two theories on more 'concrete' grounds. Explanatory and causal constraints were considered and applied to both theories under the assumption that the theory that could give some kind of causal or explanatory role to phenomenal character would be the best theory. Both theories were found to face problems on causal or explanatory grounds.

The more substantive conclusions that arose out of part one were that a semantic style of representationism does not do justice to phenomenal character; narrow content theories of content fail, and teleological accounts of content face substantial problems accounting for phenomenal content. However, Inverted Earth fails to establish the falsity of representationism.

Semantic theories fail because it is not clear how linguistic roles could have a substantial effect on the phenomenal character one undergoes when one inserts colour inverting lenses and joins a distinct linguistic community. Narrow content theories fail because it is possible to construct a scenario whereby distinct narrow contents result in shared phenomenal content or character. Teleological accounts have problem in accounting for determinate phenomenal contents.

Various replies to Inverted Earth were considered problematic. Lalor's attempt to demonstrate phenomenal re-inversion on empirical grounds was problematic. Unnoticed phenomenal re-inversions are highly unlikely. Lycan's attempt to ground unnoticed phenomenal re-inversion on the failure of transitivity is irrelevant. Lycan, Lalor and Tye's appeal to failure of memory is *ad hoc.* Finally unnoticed phenomenal re-inversion is problematic because of the likelihood of converging discriminatory capabilities in a perceiver as distinct colours follow their paths to re-inversion. This was demonstrated by considering an individual undergoing a phenomenal reinversion whilst looking at a patch of purple next to a patch of yellow.

Inverted Earth fails because of an ambiguity exploited between the lenses being part of the Earthling or part of the environment. If the lenses are part of the environment, then there is no problem with the intentional content of the Earthling's mental states staying constant. The Earthling never shares an environment with Invertlings; furthermore the lens scenario is analogous to the painted, bubbled or goggled scenarios that present no problem to the representationist. If the lenses are part of the Earthling the argument exploits an ambiguity concerning the phenomenal constancy of the Earthling's mental states. There is a clear sense in which the phenomenal content or character of the Earthling changes and the argument does not adequately acknowledge this. This is easily demonstrated by considering what an Earth object would look like to the Earthling if it were brought to Inverted Earth or if the Earthling was to return to Earth.

The more substantive conclusions of part two were that wide representationism has a serious challenge to face in giving wide contents a

causal explanatory role. The wide representationist needs to establish a solid distinction between strict and loose relations. A variety of pragmatic constraints on strict relations having an explanatory role were considered. Loose relations meet explanatory constraints but arguably the deepest explanation is given by appealing to the variables associated with the terms of the loose relation or underlying causal mechanisms. A specific reply was given two examples in the literature that were presented as an attempt to demonstrate the causal relevance of relational properties; however, both were found inadequate. The failure of Yablo's example of the photograph was particularly revealing, and Wilson's example of the relative adaptive abilities of distinct species offered a poor generalization that may well have pragmatic use but hardly offers a deep explanation.

If phenomenism appeals to non-physical intrinsic properties, then it has an even more serious problem giving phenomenal character an explanatory role. If it appeals to intrinsic non-relational physical properties, then it either has to face problems that confront type-identity theories; or if it attempts to offer a variant on functionalism with token identity and perhaps mere supervenience, then it has to account for how second order functional properties have a causal role to play.

In trying to give some kind of physicalist response to problems concerning 'explanatory gaps' the phenomenists and the representationists of a physicalist type might well stand united. It is not clear that the representationist has an advantage in overcoming 'explanatory gaps' and they still have to contend with problems concerning the causal efficacy of relational properties. The best bet for the phenomenist is to opt for some kind of physicalist account if they want to give phenomenal character a causal or explanatory role.

7.2 Future strategies for the representationists

As for a possible future direction for the representationists, I would suggest that they attempt to abandon any account of narrow content as constituting phenomenal content or character. The main problem the wide representationists confront is clarifying the nature of the relational properties

to which they appeal. I suggest that an explicit appeal to loose relations offers a better approach, but this to a large degree undermines the failure of local supervenience that an appeal to strict relations entail. The loose relations allow for no constraints on the changes in the variables associated with the terms that constitute the relation. The appeal to biological examples are often misguided and the attempted demonstration of how relations cause often rely on complex indirect causal chains that can arguable do little for the wide representationist in terms of giving them a direct explanation of the relation between phenomenal character and intentional content. I would also suggest that going general in terms of explanation is often a poor strategy in that it results in poor explanations that ignore the deep explanatory power gained by meeting basic methodological constraints of explanation.

Some of the hostility directed towards the representationist is perhaps motivated out of a feeling that they are abandoning or attempting to change basic presuppositions concerning explanatory practice. This is perhaps most clearly shown by Yablo's example where he attempts to explain why a child responds to various photographs; his inferred best explanation of the child's behaviour is because of the intentional relation the picture stands in to the child's mother. This might be a practical suggestion at the common-sense level but it would be terrible science to accept this as a deep explanation for the child's behaviour. Surely useful experiments could be done by changing variables associated with faces to detect how far and in what way shared intrinsic properties of a variety of photographs can be changed and how these changes relate to the child's response.

I suggest that the wide representationist take more care in their selection of examples if they want to offer examples that clearly demonstrate how relational properties might serve a deep explanatory role. Wilson's example of relative speciesisation is better than Yablo's, but it arguably relies on complex indirect causal chains that are perhaps not available to the representationist who wishes to explain the direct phenomenal content of mental states. I would contend that the anti-representationist has deep and substantial methodological objections to the wide-representationist approach. It is only in understanding how deep these methodological

motivations run that the wide representationist will be able to come up with an adequate explanatory account for wide content.

7.3 Future strategies for the phenomenist

As suggested earlier the phenomenist's best argument against the wide representationist is a causal or explanatory objection so it makes considerable sense for the phenomenist to accept some physicalist variant of phenomenism. The phenomenist then has to confront the 'explanatory gap', but then so does the representationist if *they* want to adopt a physicalist account.

The physicalist phenomenist also has to offer some theory that maintains an identity theory between phenomenal character and intrinsic physical properties that accounts for strong intuitions concerning multiple-realizability. Type-type identity theories have problems dealing with multiple realizability, but best offer a deep explanatory role for phenomenal character. I have already suggested that arguably the phenomenist's option is to attempt to limit the realizers that a causal function can take; furthermore, this emphasises some of the methodological explanatory constraints that were leveled against the wide representationists. If a variety of intrinsic types of physical properties realize a functional role the phenomenist should focus on the shared intrinsic properties that a variety of realizers share.

Finally one last suggestion is to speculate a physically realized functional role for phenomenal character that explains why the two most plausible theories of mind face deep causal and explanatory problems. This strategy turns most of the argumentation of the second part of this thesis on its head. Instead of trying to justify the two most plausible accounts of the relation between phenomenal character and content, one might suggest that it is because of the apparent epiphenomenal nature of phenomenal character that these theories of phenomenal character gain their plausibility. If the phenomenal character of mental states appears epiphenomenal, this explains Jackson's willingness to endorse an 'explanatory gap' and the epiphenomenality of qualia. If phenomenal character appears

epiphenomenal, this explains the intuitive plausibility of the inverted spectrum scenarios that motivate the phenomenist position. If phenomenal character appears epiphenomenal, this perhaps partly explains why the representationist latches onto relational properties to attempt to explain content. As argued in chapters four and five it is a feature of relations that they tend to allow no role for deep explanation, or if they explain, it is only in terms of the variables associated with the terms that constitute the relation. If phenomenal character actually has a role that gives it the appearanceof epiphenomenality, then this not only explains why it is so hard to give phenomenal charactera causal or explanatory role, it explains why two of the most plausible theories of phenomenal character face deep causal or explanatory problems.

This account does not support an *actual* epiphenomenalism of any type. It merely attempts to explain the *appearance* of epiphenomenality. The causal or explanatory problems are not in fact an entailment of the two accounts, but indeed essential to the two accounts. Once the apparent epiphenomenality of phenomenal characteris explained then a cogent theory of the relation between phenomenal characterand content might be offered.

7.4 Final speculation

One possible role for phenomenal character that explains its apparent epiphenomenality could be a role that gives phenomenal character an inhibitory function. If the intrinsic properties that realize this functional role serve to block responses to stimuli, then this arguably explains the appearanceof epiphenomenality. If one dwells on the phenomenal character of a thought, one might realize that the phenomenal characterarises out of the words that are not spoken. The content of the thought arises from the content of the words that would have normally constituted the speech act that never occurred. If one hums a tune in one's head, the content of the internalized mental state is the song that is never actualized. The phenomenal character arises from the inhibition of the action that never materializes. Why does one not hear another person's thoughts? The process that blocks the spoken words that would normally arise from that mental state constitutes both the content and the phenomenal character of

that person's thoughts. From a third person perspective it is notoriously hard to see an internal process that by its very nature blocks the only means to gaining access to that process.

Pain presents a particular challenge to this theory because it is usually taken to be the paradigm case of a phenomenal state that clearly has a somewhat direct causal role in *promoting* action. One solution is to suggest that there are various pathways a stimulus can take to provoke a response, perhaps a cognitive blocking occurs that send the pain down a more direct pathway thus provoking a direct response. Alternatively what should be emphasized is the possibility of having substantial damage inflicted on oneself yet one need not respond automatically to the damage given a greater priority.

Perceptual experiences present a challenge to this theory because there is no clear action being blocked by the myriad of visual input that one receives just by looking around the room. In the case of perceptual experience perhaps the cognitive blocking is occurring not just before the output, but also at the end of the input processing. If all the visual information is at some point simultaneously cognitively blocked at the end of the processing, this then explains the phenomenal character of the perceptual experience. Cognitive blocking could have a variety of roles to play, at the end of an input or just before an output, or to send a stimulus down a distinct causal pathway, but whatever the stage in the process, it is the inhibition that accounts for the phenomenal. The content is either explained by the information contained by the stimulus or the content of the overt action blocked by the cognitive inhibition.

Bibliography

Antony, L., & Levine, J. (1997). Reduction with autonomy. *Philosophical Perspectives*11, 83-106.

Armstrong, D. M., & Malcolm, N. (1984). *Consciousness and Causality.* Oxford: Blackwell.

Akins, K. (1993). What is it like to be boring and myopic? In B. Dahlbom (Ed.). *Dennett and His Critics.* Oxford: Blackwell.

Block, N. (1990). Inverted Earth. *Philosophical Perspectives*4, 51-79.

Block, N. (1996). Mental paint and mental latex. In E. Villanueva (Ed.). *Perception, Philosophical Issues 7* (pp. 19-49). Ridgeview: Atasadero.

Block, N. (1997). Anti-Reductionism slaps back. *Philosophical Perspective* 11, 107-32.

Block, N. (1999).Sexism, ageism and racism and the nature of consciousness. *Philosophical Topics* 26, 39-69.

Braddon-Mitchell, D. & Jackson, F. (1997). The teleological Theory of Content. *Australasian Journal of Philosophy* 75, 474-89.

Burge, T. (1986). Individualism and Psychology. *Philosophical Review* 95, 3-45.

Burge, T. (1995). Reply: Intentional properties and causation. In C. Macdonald, & G. Macdonald (Eds.). *Philosophy of Psychology: Debates on Psychological Explanation* (Vol 1). Cambridge, MA: Blackwell.

Burge, T. (1996). Two kinds of consciousness. In N. Block, & O. Flanagan, & G. Güzeldere (Eds.). *The Nature of Consciousness: Philosophical and Scientific Debates.* Cambridge, MA: MIT Press.

Byrne, A., & Hilbert, D. (1997). Colors and reflectances. In A. Byrne, & D. Hilbert (Eds.). *Readings on Color: The Philosophy of Color* (Vol 1). Cambridge, MA: MIT Press.

Chalmers, D.J. (1996). *The Conscious Mind.* New York: Oxford University Press.

Churchland, P.M. (1985). Reduction, Qualia, and the direct introspection of brain states. *Journal of Philosophy* 82, 8-28.

Clark, A. (1985). Spectrum inversion and the color Solid. *The Southern Journal of Philosophy* 23, 431-43.

Crane, T. (1991). All the difference in the world. *The Philosophical*

Quarterly 41, 1-25.

Dolezal, H. (1982). *Living in a World Transformed - Perceptual and Performatory Adaptation to Visual Distortion.* New York: Academic Press.

Dennett, D.C. (1978). Why you can't make a computer that feels pain. In D. C. Dennett. *Brainstorms*. Cambridge, MA: Bradford Books.

Dennett, D.C. (1988). Quining qualia. In A. Marcel, & E. Bisiach (Eds.). *Consciousness in Contemporary Science.* Oxford: Oxford University Press.

Dennett, D.C. (1991). *Consciousness Explained.* Boston, MA: Little, Brown & Co.

Dretske, F. (1981). *Knowledge and the Flow of Information* . Cambridge, MA: MIT Press.

Dretske, F. (1990). Reply to reviewers. *Philosophy and Phenomenaological Research* 50, 819-40.

Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, MA: MIT Press.

Fodor, J. (1974). Special sciences. *Synthese* 28, 97-115.

Fodor, J. (1987). *Psychosemantics.* Cambridge, MA: MIT Press.

Fodor, J. (1990). Banish DisContent. In W. Lycan (Ed.). *Mind and cognition.* Oxford: Basil Blackwell.

Fodor, J. (1994). A theory of narrow content II. In S. Stich, & T. Warfield (Eds.). *Mental Representation: A Reader.* Oxford: Blackwell.

Fodor, J. (1995). A modal argument for narrow content. In C. Macdonald, & G. Macdonald (Eds.). *Philosophy of Psychology: Debates on Psychological Explanation* (Vol 1). Oxford & Cambridge, MA: Blackwell.

Gert, B. (1965). Imagination and Verifiability. *Philosophical Studies* 16, 44-47.

Goodman, N. (1977). *The Structure of Appearance* (3rd ed.). Boston: Dordrecht Reidel.

Hardin, C. L. (1988). *Color for Philosophers: Unweaving the Rainbow*. Indianapolis: Hacket.

Harman, G. (1990). The intrinsic quality of experience. In *Philosophical Perspectives* 4, 31-52.

Held, R., & Blossom, J. (1961). Neonatal deprivation and adult rearrangement:Complementary techniques for analyzing plastic

sensory-motor coordinations. *Journal of Comparative and Physiological Psychology* 54, 33-7.

Hempel, C. (1965). Aspects of scientific explanation. In C. Hempel (Ed.). *Aspects of scientific explanation and other essays in the philosophy of science* (pp. 331-496). New York: Free Press.

Hempel, C., & Oppenheim, P. (1965). Studies in the Logic of Explanation. In C. Hempel (Ed.). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science* (pp. 245-295). New York: Free Press.

Horgan, T. (1997).Kim on mental causation and causal exclusion. *Philosophical Perspectives* 11, 165-164.

Jackson, F. (1982). Epiphenomenal Qualia. *Philosophical Quarterly*, 32,127-36.

Kim, J. (1992). Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research* 52, 1-26.

Kim, J. (1993). The Non-Reductivists troubles with Mental Causation. In J. Hele, & A. Mele (Eds.). *Mental Causation*. Oxford: Oxford University Press.

Kim, J. (1997). The Mind-Body Problem. *Philosophical Perspectives* 11, 185-207.

Lalor, B. (1999). Intentionality and Qualia. *Synthese* 121, 249-89.

Levine, J. (1983). Materialism and Qualia: The explanatory gap. *Pacific Philosophical Quarterly* 64, 354-61.

Levine, J. (1997). Are qualia just representations? *Mind and Language* 12, 101-13.

Lewis, D. (1988). What Experience Teaches. In W. Lycan (Ed.).*Mind and Cognition: A Reader* (pp. 499-519). Oxford: Basil Blackwell.

Loar, B. (1990).Phenomenal states. *PhilosophicalPerspectives*4, 81-108.

Locke, J. (1690). *An Essay Concerning Human Understanding.* (A.C. Fraser ed., 1959). New York: Dover.

Lombard, B. (1999). Events. In S. Hales (Ed.). *Metaphysics* (pp.348-370). London: Wadsworth.

Lycan, W. (1973). Inverted Spectrum. *Ratio* 15, 315-19.

Lycan, W. (1996). *Consciousness and experience* Cambridge, MA: MIT Press.

McGinn, C. (1986). *Mental Content.* Oxford: Blackwell.

McGinn, C. (1991). *The Problem of Consciousness.* Oxford: Blackwell.

Millikan, R. G. (1993). *White Queen Psychology and Other Essays for Alice.* Cambridge, MA: MIT Press.

Moore, G.E. (1922). The refutation of Idealism. In G. E. Moore. *Philosophical Studies*. London: Routledge and Kegan Paul, Ltd.

Neander, K. (1995). Misrepresenting and malfunctioning.*Philosophical Studies* 79, 109-41.

Nemirow, L. (1990). Physicalism and the cognitive role of acquaintance. In W. Lycan (Ed.). *Mind and Cognition: A Reader* (490-499). Oxford: Basil Blackwell.

Papineau, D. (1990). Truth and teleology. In D. Knowles (Ed.). *Explanation and its Limits*. Cambridge: Cambridge University Press.

Papineau, D. (1997.). Teleosemantics and indeterminacy. *Australasian Journal of Philosophy* 76, 1-14.

Papineau, D. (Forthcoming). *Thinking about consciousness.*

Peacocke, C. (1983). *Sense and Content.* Oxford: Oxford University Press.

Perry, J. (1979). The problem of the essential indexical. *Nous* 13, 3-21.

Putnam, H. (1975a). The meaning of 'meaning'. In H. Putnam. *Mind Language and reality.* Cambridge: Cambridge University Press.

Putnam, H. (1975b). Philosophy and our mental life. In H. Putnam. *Mind Language and reality.* Cambridge: Cambridge University Press.

Putnam, H. (1975c). The nature of mental states. In H. Putnam. *Mind Language and reality.* Cambridge: Cambridge University Press.

Quine, W.V. (1969). *Ontological relativity and other essays.* New York: Columbia University Press.

Rey, G. (1998). A narrow representationalist account of qualitative experience.*Philosophical Perspectives* 12, 435-57.

Salmon, W. S. (1999). *Causality and Explanation.* Oxford: Oxford University Press.

Shoemaker, S. (1975). Functionalism and Qualia. *Philosophical Studies* 27, 291-315.

Shoemaker, S. (1982). The Inverted Spectrum. *Journal of Philosophy* 79, 357-81.

Shoemaker, S. (1998). Two cheers for Representationism. *Philosophy and Phenomenological Research* 3, 663-78.

Stalnaker, R. (1999). Comparing Qualia across Persons. *Philosophical Topics* 26, 385-405.

Stich, S. (1990). Building belief: Some queries about representation, indication, and function. *Philosophy and Phenomenological Research* 50, 801-6.

Taylor, D. M. (1966). The incommunicability of content. *Mind* 75, 527-41.

Tye, M. (1992). Visual qualia and visual content. In T. Crane (Ed.). *The Contents of Experience.* Cambridge: Cambridge University Press.

Tye, M. (1994). Qualia content and the inverted spectrum. *Nous* 28, 159-83.

Tye, M. (1995a). Blindsight, orgasm and representational overlap. *Behavioral and Brain Sciences* 18, 268-69.

Tye, M. (1995b). *Ten Problems of Consciousness*. Cambridge: MIT Press.

Tye, M. (1998a). Response to discussants. *Philosophy and Phenomenological Research* 3, 679-87.

Tye, M. (1998b). Inverted Earth, Swampman, and Representationism. *Philosophical Perspectives* 12, 459-78.

Tye, M. (1999). Phenomenal consciousness: the explanatory gap as cognitive illusion. *Mind*, 108, 705-725.

Van Gulick, R. (1995). How should we understand the relation between intentionality and phenomenal consciousness. *Philosophical Perspectives* 9, 271-89.

Walsh, D. (1998). Wide Content Individualism. *Mind* 107, 625-50.

Wilson, R. (1997). *Cartesian Psychology and Physical Minds.* Cambridge: Cambridge University Press.

Wittgenstein, L. (1958). *Philosophical Investigations.* New Jersey: Prentice Hall.

Wittgenstein, L. (1980). *Remarks on the Philosophy of Psychology* (Vols. 1 -2). Chicago: The University of Chicago Press.

Yablo, S. (1997). Wide Causation. *Philosophical Perspectives* 11, 251-82.